

Der Umgang mit Variablen bei offenen Experimentieraufgaben im Physikunterricht

Eine Beobachtungsstudie am Beispiel der Konstruktion von
auftriebserzeugenden Profilen für ein Windradmodell

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von
Stefan Kirchner

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Stefan Hecht, Ph.D.

Gutachter:

1. Prof. Dr. Burkhard Priemer
2. Prof. Dr. Rita Wodzinski
3. Prof. Dr. Lutz-Helmut Schön

Tag der mündlichen Prüfung: 11. Februar 2013

Zusammenfassung

Diese Forschungsarbeit beschreibt und bewertet beobachtbare Experimentierprozesse hinsichtlich der grundlegenden naturwissenschaftlichen Arbeitsweise „Umgang mit Variablen“, die Gymnasiasten am Ende der Sekundarstufe I bei der Bearbeitung einer offenen Experimentieraufgabe im Physikunterricht zeigen. Dargelegt wird zudem der Nutzen einer differenzierten Begriffsbestimmung für eine zielorientierte und strukturierte Entwicklung von offenen Experimentieraufgaben.

In einer empirisch-quantitativen Exploration mit 82 Schülern wird eine offene Aufgabenstellung zum Thema Windenergie eingesetzt, die den Experimentierenden bezüglich der naturwissenschaftlichen Arbeitsweise einen hohen Grad an Entscheidungsmöglichkeiten erlaubt. Die Studie möchte für den Umgang mit Variablen die Wirksamkeit der offenen Experimentieraufgabe aufzeigen und zwar unter der Bedingung, dass Schüler vor der Bearbeitung der offenen Experimentieraufgabe eine vorbereitende bzw. keine vorbereitende Instruktion erhalten. Das Hauptinteresse der vergleichenden Studie liegt bei einem Kontexttransfer, den die Schüler von der vorbereitenden Instruktion auf die offene Experimentieraufgabe leisten müssen, wenn sie a) keinen Wechsel der Experimentierbedingungen und b) einen Wechsel der Experimentierbedingungen vorfinden.

Die Untersuchung beruht auf einem Kontrollgruppendesign, in dem die Experimentierprozesse mittels eines Beobachtungsbogens erfasst werden. Das erhobene Datenmaterial stellt die Basis für eine kriteriengeleitete Bewertung und Einweisung der Experimentierprozesse in Niveaustufen dar. Um die qualitativen Ergebnisse zu stützen, werden statistische Verfahren eingesetzt.

Die Ergebnisse zeigen, dass die Fähigkeit der Probanden mit Variablen umzugehen, durch den Einfluss einer vorbereitenden Instruktion erheblich gesteigert wird. Eine Abhängigkeit vom Grad der Kontextunterschiede zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe wird nur gefunden, wenn die Probanden neben den veränderten Kontexten auch veränderte Experimentierbedingungen vorfinden. Die Ergebnisse verleiten für die Vermittlung der naturwissenschaftlichen Arbeitsweise „Umgang mit Variablen“ zu der forschungsbezogenen Aussage: Mehr Offenheit beim Experimentieren wagen!

Abstract

This research paper describes and rates observable experimental processes secondary school students' (aged 16) reveal concerning the fundamental scientific method of using variables in open-ended experimental tasks in the physics classroom. In addition, the benefit of a corresponding sophisticated definition will be shown in order to enable a goal-oriented and structured development of open-ended experimental tasks.

In an empirical quantitative exploration with 82 school students, an open-ended problem on the topic “wind power” is utilized, allowing a high degree of autonomy in decision-making concerning the scientific method of using variables. The study aims at demonstrating the effectiveness of open-ended experimental tasks depending on whether the students received preparing instruction beforehand. The main interest of the comparative study lies in a transfer of context that students need to perform between the preparing instruction and the open-ended experimental task with or without a shift of experimental conditions.

The investigation is based on a control-group design, capturing experimental processes by use of a monitoring sheet. The collected data forms the basis for a criteria-oriented rating of experimental processes and their assignment to certain levels. The qualitative results are supported by the usage of statistical methods.

The results show that the students' ability to use variables is highly enhanced by receiving preparing instructions. Dependence on the degree of context-shift between preparing instruction and open-ended experimental task is only the case when students find beyond shifted contexts altered experimental conditions. Concerning the teaching of the scientific method of using variables, the results lead to a research-related statement: dare to employ more open-ended experiments!

Inhaltsverzeichnis

1	Einleitung	1
	Theoretischer Teil	3
2	Offenes Experimentieren und Naturwissenschaftliche Arbeitsweisen	5
2.1	Offenes Experimentieren.	5
2.1.1	Divergente Begriffsbestimmungen und Bewertungen des offenen Experimentierens.	6
2.1.2	Eigener Ansatz des offenen Experimentierens.	12
2.1.3	Perspektiven des offenen Experimentierens	19
2.2	Naturwissenschaftliche Arbeitsweisen	21
2.2.1	Prozessbezogene Fähigkeiten des Experimentierens	22
2.2.2	Der Umgang mit Variablen beim Experimentieren.	25
2.3	Entwicklung eines Aufgabenbeispiels zum offenen Experimentieren.	29
2.3.1	Kriterien zur Entwicklung und Konstruktion des Aufgabenbeispiels.	29
2.3.2	Präsentation und Diskussion des entwickelten Aufgabenbeispiels	32
2.3.3	Verortung des Aufgabenbeispiels im eigenen Ansatz des offenen Experimentierens.	36
2.3.4	Kritische Anmerkungen zur Entwicklung des Aufgabenbeispiels.	39
	Empirischer Teil	41
3	Explorative Vorstudien	43

3.1	Explorative Vorstudie I.	44
3.2	Konsequenzen der ersten Vorstudie.	49
3.3	Explorative Vorstudie II.	53
3.4	Ergänzungen und kritische Anmerkungen	56
3.5	Ausblick Hauptstudie	58
4	Hauptstudie	65
4.1	Forschungsfragen.	65
4.2	Methoden	66
4.2.1	Design, Stichprobe, Teilstichprobe, Variablen.	67
4.2.2	Messinstrument Fach- & NAW-Test.	72
4.2.3	Messinstrument Beobachtungsbogen	81
4.3	Ergebnisse des Fach- & NAW-Tests	101
4.4	Ergebnisse der Beobachtung.	112
4.4.1	Prozessdiagramme und Prozessanalyse des Experimentierens.	112
4.4.2	Kriterien zur Bewertung und Einweisung der Experimentierprozesse in Niveaustufen.	122
4.4.3	Experimentiermuster und Experimentiertypen.	124
4.4.4	Niveaustufen des Experimentierens.	133
4.5	Korrelation und Triangulation der Ergebnisse: NAW-Test & Beobachtung.	147
5	Hypothesen	159
6	Zusammenfassung und Ausblick	161
	Anhang	171
	Abbildungsverzeichnis	229
	Tabellenverzeichnis	231
	Literaturverzeichnis	235

1 Einleitung

PISA, NBS, KP. Diese und weitere Akronyme sind geeignet, die jüngsten Wandlungsprozesse im Bildungssystem zu erfassen, die im Kontext von Schule und Unterrichtswissenschaften für spürbare Veränderungen gesorgt haben. Zwei Kennzeichen, die für den „Paradigmenwechsel“ stehen, sind, dass internationale Schulvergleichsstudien wie das Programme for International Student Assessment (PISA) Literacy Konzepten folgen und dass Unterricht durch überprüfbare Output-Variablen gesteuert wird. Die in den Nationalen Bildungsstandards (NBS) und den landesspezifischen Kernlehrplänen (KP) aufgestellten Standards sind normativ gesetzt worden. Daher ist es möglich, dass sie den empirisch beschreibbaren Fähigkeiten der Lernenden nicht gerecht werden.

Ein zentrales Anliegen dieser Studie ist es hingegen, die Fähigkeiten von Probanden an der Schnittstelle zur gymnasialen Oberstufe deskriptiv zu erfassen. Dieses Ziel wird für den Umgang mit Variablen beim Experimentieren verfolgt, der ein Beispiel für eine grundlegende naturwissenschaftliche Arbeitsweise und damit wesentlicher Bestandteil des Physikunterrichts ist. Gemessen an der Bedeutung des Themas überrascht es, dass aus den naturwissenschaftsdidaktischen Disziplinen kaum Erkenntnisse für den Umgang mit Variablen bekannt sind, die auf direkten Beobachtungen beim Experimentieren fußen. Stattdessen werden häufig aus test-ökonomischen Gründen paper-pencil-basierte Messinstrumente als Ersatzmaß für die direkte Beobachtung von Experimentierprozessen eingesetzt. Kritisch zu hinterfragen ist, ob mit Ersatzmaßen dieselben experimentellen Fähigkeiten gemessen und abgebildet werden können. Ergebnisse mehrerer Korrelationsanalysen weisen darauf hin, dass die Passung zwischen der direkten Beobachtung von Experimenten (Referenzmaß) und den paper-pencil-basierten Testinstrumenten (Substitutionsmaß) nur unzureichend gelingt. Um die in den Nationalen Bildungsstandards geforderten prozessbezogenen Fähigkeiten beim Experimentieren dennoch präziser erfassen zu können, werden von den Didaktiken der Naturwissenschaften und der Lehr-Lern-Forschung Alternativen zu den schriftlichen Messverfahren erforscht. Hierzu gehört beispielsweise der Einsatz von computerbasierten Lernumgebungen.

Entgegen dieser Forschungsrichtung wird in der vorliegenden Studie auf den Gebrauch von Substitutionsmaßen verzichtet, indem die Experimentierprozesse von Probanden, die sie in einer realen Experimentierumgebung für den Umgang mit Variablen zeigen, kriteriengeleitet beobachtet, analysiert und bewertet werden. Hierbei wird zwischen den allgemeinen Fähigkeiten im Umgang mit Variablen und den kontextspezifischen Fähigkeiten beim Experimentieren unterschieden. Zur Aufnahme der Experimentierprozesse wird eine offene Aufgabenstellung gewählt, die nach dem eigenen Ansatz des offenen Experimentierens entwickelt und am Beispiel der Konstruktion von Profilen für das Modell eines Windrades realisiert wird. Um möglichen Schwierigkeiten wie Verstößen gegen die Variablenkontrollstrategie bei der Planung, Durchführung und Auswertung von Experimenten zu begegnen, wird ein didaktischer Ansatz verfolgt, der den Schwerpunkt auf eine informierende, explizite und ganzheitliche Vorbereitung von physikalischen Arbeitsweisen setzt. Die Umsetzung des didaktischen Ansatzes umfasst ein Konzept mit mehreren Ausprägungen, die in Bezug auf die offene Experimentieraufgabe unterschiedliche Transferleistungen erfordern.

Eine Besonderheit der entwickelten offenen Experimentieraufgabe liegt darin, dass mit ihrem Einsatz nur ein und nicht mehrere Hauptziele gleichzeitig verfolgt werden. Hierbei wird Abstand von dem Ansatz genommen, innerhalb einer Experimentieraufgabe sowohl hohe fachinhaltliche Anforderungen und schwierige Methoden der Erkenntnisgewinnung als auch anspruchsvolle experimentelle Fertigkeiten vereinen zu wollen, so dass eine Anhäufung von verschiedenen Schwierigkeiten (inhaltliches Wissen, strategisches Wissen, methodisches Wissen) vermieden wird. Die Schwerpunktsetzung auf die naturwissenschaftlichen Arbeitsweisen am Beispiel des Umgangs mit Variablen beim Experimentieren wird als Antwort auf die in der Literatur kritisch beschriebene Vielfalt von Zielen verstanden, welche Lernende beim eigenständigen Experimentieren zu erfüllen haben und die die Gefahr einer möglichen Überforderung bergen. Zu dieser Einschätzung kommt auch Hodson (1993), wenn er das Verhalten der Experimentierenden mit „very busy getting nowhere“ beschreibt und folgert, dass sich nicht alle Zielsetzungen von Schülerexperimenten in einem experimentellen Ansatz gleichzeitig verwirklichen lassen.

Sowohl die Schwerpunktsetzung auf naturwissenschaftliche Arbeitsweisen als auch der vorbereitende didaktische Ansatz zum Experimentieren werden in dieser Studie als Unterstützungsmaßnahmen eingesetzt, die verhindern sollen, dass die Konstruktion von Windradprofilen wie bei Miguel de Cervantes Saavedras Don Quijote im „Kampf gegen Windmühlen“ endet.

Theoretischer Teil

2 Offenes Experimentieren & Naturwissenschaftliche Arbeitsweisen

Experimente respektive Schülerexperimente werden seit nahezu 200 Jahren in der Lehre über die Naturwissenschaften eingesetzt (Edgeworth & Edgeworth, 1811). Entsprechend der frühen Nutzung existiert eine umfangreiche pädagogische und (fach-)didaktische Forschungstradition. Allerdings werden hieraus für den theoretischen Teil der Arbeit nur diejenigen didaktischen und methodischen Aspekte des Experimentierens ausgewählt, welche mit dem Arbeitstitel abgestimmt werden können. Dazu gehört u. a. ein Bericht zum Stand der Forschung über das offene Experimentieren, über die prozessbezogenen Kompetenzen beim Experimentieren und über die naturwissenschaftlichen Arbeitsweisen, vor allem für den Umgang mit Variablen. Weiterhin wird im Theorieteil das für die Vor- und Hauptstudien relevante, selbst entwickelte Aufgabenbeispiel (Kap. 2.3) aufgenommen, welches auf den fachspezifischen Aspekten des Experimentierens zum Umgang mit Variablen (Kap. 2.2) und auf dem eigenen Ansatz des offenen Experimentierens (Kap. 2.1) fußt. Obwohl die lange Forschungstradition des Experimentierens mehr als die benannten Aspekte umfasst, möchte die vorliegende Arbeit einer allgemeinen theoretischen Abhandlung des Experimentierens nicht nachkommen. Für einen Überblick sei beispielhaft auf Arbeiten von Hodson (1992), Leach & Paulsen, (1999), Tiberghien, Veillard, Le Maréchal, Buty & Millar (2001), Psillos & Niedderer (2002), Séré (2002), Lunetta, Hofstein & Clough (2007) verwiesen.

2.1 Offenes Experimentieren

Der Theorieteil zum offenen Experimentieren wird durch drei Unterkapitel strukturiert. Im Kapitel 2.1.1 wird über den Forschungsstand zum offenen Experimentieren berichtet. Der Vergleich zahlreicher Begriffsbestimmungen führt

zu dem Entschluss, einen eigenen Ansatz des offenen Experimentierens zu entwickeln, der im Kapitel 2.1.2 vorgestellt wird. Die sich hieraus ergebenden Perspektiven des offenen Experimentierens werden im Kapitel 2.1.3 diskutiert.

2.1.1 Divergente Begriffsbestimmungen und Bewertungen des offenen Experimentierens

Der Begriff des offenen Experimentierens (engl. *open ended*, *open inquiry*, *inquiry based experiments*, vgl. Lunetta, Hofstein & Clough, 2007) zeichnet sich im fachdidaktischen Sprachgebrauch dadurch aus, dass er häufig in Verbindung mit begriffserweiternden Attributen (Anlagerungen) verwendet wird. Beispielsweise werden zur Charakterisierung des offenen Experimentierens von Reinhold (1996, S. 87) das Attribut selbständig „[...] das offene, relativ selbständige Experimentieren“, von Fischer & Draxler (2001, S. 390), Trendel & Fischer (2007, S. 393) die Anlagerungen nicht strukturiert „[...] eine offene, nicht strukturierte Aufgabenstellung“ und von Mikelskis-Seifert & Rabe (2007, S. 37) das Attribut problemorientiert „[...] das offene, problemorientierte Experimentieren“ benutzt. Die begriffserweiternden Zusätze lassen erahnen, dass verschiedene Konzepte und Auffassungen des offenen Experimentierens existieren. Dieser Eindruck erhärtet sich beim Vergleich ausgewählter Ansätze. Um die zahlreichen Begriffsbestimmungen sinnvoll gegenüberzustellen, wird zunächst eine Sortierung eingeführt, die zwischen phänomenorientierten Ansätzen des offenen Experimentierens und Ansätzen mit festgelegten Öffnungsgraden unterscheidet (s.u.). Die Anzahl der Ansätze, die in den Rubriken benannt werden, spiegeln den bei der Literaturarbeit gewonnenen Eindruck wieder, dass in den letzten Jahrzehnten vermehrt Konzepte mit festgelegten Öffnungsgraden erstellt wurden. Weiterhin kann eine Minderheit von Begriffsbestimmungen identifiziert werden, die weder der einen noch der anderen Rubrik zuzuordnen sind (vgl. Fischer & Draxler, 2001; Trendel & Fischer, 2007; Engeln, 2004; Hopf, 2007, s. u.). Sofern die Begriffsbeschreibungen es erlauben, werden auch diese Konzepte und Auffassungen im eigenen Ansatz des offenen Experimentierens verortet (vgl. Priemer, 2011). Die Verortung trägt dazu bei, dass die zahlreichen Begriffsbestimmungen des offenen Experimentierens systematischer verglichen werden können. Ergänzend sei hinzugefügt, dass die Vielfalt und Multidimensionalität der Ansätze nicht nur auf das offene Experimentieren beschränkt sind. Ein analoger Sachverhalt liegt auch bei den Begriffsbestimmungen des offenen Unterrichts vor (Heinzel, 2006). Dennoch gelingt es, dem offenen Unterricht Dimensionen zuzuweisen, die eine klare Trennung zur Begriffsbestimmung des offenen Experimentierens markieren. Als Beispiel seien Dimensionen wie organisatorische, soziale und persönliche Offenheit benannt, die aus der Definition des offenen Unterrichts von Peschel (2002, S. 78)

herausgefiltert werden können. Nichtsdestotrotz lassen einige der präsentierten Ansätze des offenen Experimentierens die notwendige Trennschärfe zum offenen Unterricht nicht erkennen (s. u.). Genauere Ausführungen zur Abgrenzung der Begriffe offenes Experimentieren, offener Unterricht und offene Aufgaben finden sich bei Priemer (2011).

Phänomenorientierte Ansätze des offenen Experimentierens. In diesem Abschnitt werden Ansätze des offenen Experimentierens zusammengefaßt, die physikalische Phänomene zum Gegenstand des offenen Experimentierens machen. Diese Variante des offenen Experimentierens wird in Arbeiten von Heege (1978), Heege & Bruns (1981), Reinhold (1996, 1997a, 1997b) und Braun (2009) vertreten. Während Reinhold und Braun eine höhere Übereinstimmung hinsichtlich ihrer Auffassungen zum offenen Experimentieren zeigen, können deutliche Unterschiede zum Ansatz von Heege bzw. Heege & Bruns benannt werden. Leider stellen die Autoren ihre Konzepte nur zum Teil explizit vor, so dass auf kontrastierende Zitate zugunsten einer beschreibenden Begriffsbestimmung verzichtet werden muss.

Reinhold (1996, 1997b), der mit seinem Ansatz einen Rückbezug zu Heege (1978) bzw. Heege & Bruns (1981) herstellt, schreibt dem offenen Experimentieren eine feste Struktur mit den Strukturelementen Erklärungsbedürftige Situation, Probieren und Spekulieren, Reflexion der gebildeten Systeme, Normalwissenschaftliches Experimentieren, Reflexion der Systembildung und Reflexion der Lerntätigkeit zu. Die Offenheit beim Experimentieren wird auf ausgesuchte Phasen (Strukturelemente) beschränkt. Sie zeigt sich insbesondere in der Phase des Probierens und Spekulierens. Das hat zur Konsequenz, dass die Aussagen zur Wirksamkeit der Fallstudien auf die entsprechenden Strukturelemente bezogen werden müssen (s. u.). In Anlehnung an Reinhold (1996) weist Braun (2009) dem offenen Experimentieren ebenfalls eine feste Struktur zu. Beide Ansätze unterscheiden sich nur geringfügig, weil Braun fast identische Strukturelemente benennt. Braun legt den Fokus jedoch auf die fachlichen Analysen der Phänomene, deren Einsatz in der universitären Lehre durch Erfahrungsberichte dokumentiert wird.

Obwohl sich Reinhold auf Heege bzw. Heege & Bruns bezieht, sind in deren Arbeiten weniger feste, durch Strukturelemente gekennzeichnete Abläufe des offenen Experimentierens zu erkennen. Vor allem aus der Quelle von Heege & Bruns (1981) kann herausgelesen werden, dass es den Autoren beim offenen Experimentieren („Improvisationen“) vorwiegend um die Konzeption von Demonstrationsexperimenten („Improvisationsexperimente, Freihandversuche“) geht, die zu ausgesuchten Phänomenen entwickelt werden sollen. Trotz thematischer Vorgaben, sind die Experimentierphasen Planung, Durchführung und Auswertung für die Lernenden offen. Weiterhin sind zur Lösung der Experimentieraufgaben auch Trial-& Error-Methoden zugelassen.

Mit einem Vorgriff auf Kapitel 2.1.2 darf schon jetzt berichtet werden, dass das Konzept von Heege bzw. Heege & Bruns im eigenen Ansatz des offenen Experimentierens verortet werden kann. Hauptsächlich aus der Arbeit von Heege & Bruns (1981) kann an verschiedenen Stellen herausgefiltert werden, dass die Autoren Aspekte des offenen Experimentierens beschreiben, die den Dimensionen *Fachinhalt-Strategie-Methode* (1. Dimension), *Lösungen und Lösungswege* (2. Dimension) und den *Phasen des Experimentierens* (3. Dimension) zugeordnet werden können. Dem eigenen Ansatz folgend, werden von Heege & Bruns keine Vorgaben zur *Strategie* und zur *Methode* gemacht und mit Ausnahme der Aufgabenstellung besteht Offenheit auch in allen Phasen des Experimentierens. Zudem werden die präsentierten Aufgabenbeispiele von mehreren Lösungen und mehreren Lösungswegen unterstützt. Dagegen sind von Reinhold Aufgabenbeispiele bekannt, die prinzipiell einem Lösungsweg (ggf. mit Variationen) folgen, der zu einer Lösung (Erklärung) führt (Reinhold, 1996, S. 169 bzw. S. 204).

Ansätze des offenen Experimentierens mit festgelegten Öffnungsgraden. In diesem Abschnitt werden Ansätze des offenen Experimentierens zusammengestellt, die auf einer gemeinsamen Herangehensweise der Begriffsbildung basieren. Es soll aufgezeigt werden, dass diese Ansätze trotz ihrer Parallelen zu divergenten Begriffsbestimmungen führen. Die Parallelen, die in der Methodik der Begriffsbestimmung liegen, können auf Schwab (1962) und Herron (1971) zurückgeführt werden. Mit einem Zitat von Schwab soll die Herangehensweise, wie offenes Experimentieren mit festgelegten Öffnungsgraden charakterisiert wird, illustriert werden:

“At the simplest level, the manual can pose problems and describe ways and means by which the student can discover relations he does not already know from his books. At a second level, problems are posed by the manual but methods as well as answers are left open. At a third level, problem, as well as answer and method, are left open: the student is confronted with the raw phenomenon – let it be even as apparently simple a thing as a pendulum.” (Schwab, 1962, S. 55)

Dem Zitat ist zu entnehmen, dass der Autor ein Set von Kategorien ‘problem, method, answer’ erstellt, aus dem Grade des Experimentierens (Öffnungsgrade) abgeleitet werden. Vor der Zuweisung der Experimentiergrade muss entschieden werden, welche Kategorien offen bzw. geschlossen sind. Die Entscheidung führt zu mehreren Kompositionen aus offenen und geschlossenen Kategorien, die von Schwab, dem Zitat folgend, in drei und von Herron (1971) in vier Öffnungsgrade überführt werden. Dieses Prinzip der Begriffsbildung wird von vielen Autoren angewendet. Als Beispiele seien die Arbeiten von Hegarty (1978), Hegarty-Hazel (1986 zitiert nach Staer, Goodrum & Hackling, 1998), Mayer & Ziemek (2006), Buck, Bretz & Towns (2008) und Hof (2010) genannt. Weitere Beispiele können bei

Staer, Goodrum & Hackling (1998) sowie bei Millar, Tiberghien & Le Maréchal (2002) entnommen werden.

Die Ansätze stimmen darin überein, dass offenes Experimentieren quantitativ erfasst wird. Die Quantifizierung wird umgesetzt, indem Öffnungsstärken auf Kategoriensets zurückgeführt und auf mehrstufigen Skalen abgebildet werden. Die Skalierungen werden in der Literatur mit Formulierungen wie *levels of openness* (Schwab, 1962) oder *degrees of openness* (Herron, 1971) gekennzeichnet. Einige Autoren weisen den Freiheitsgraden des Experimentierens sogar Instruktions- bzw. Experimentierstile zu, so dass die quantifizierten Öffnungsgrade in *levels of inquiry (enquiry)* überführt werden. Als Beispiele seien die fünfstufigen Skalen von Hegarty-Hazel (1986 zitiert nach Staer et al., 1998) mit den Skalenstufen 0, 1, 2a, 2b, 3 und von Buck, Bretz & Towns (2008) mit den Skalenstufen 0, ½, 1, 2, 3 angeführt. Nach Hegarty-Hazel werden den Freiheitsgraden folgende Instruktionsarten zugeordnet: level 0 (verification), level 1 (guided inquiry), level 2 a/ b (open guided inquiry), level 3 (open inquiry). Nach Buck et al. (2008) werden den Öffnungsgraden des Experimentierens folgende Instruktionsarten zugewiesen: level 0 (confirmation), level ½ (structured inquiry), level 1 (guided inquiry), level 2 (open inquiry), level 3 (authentic inquiry).

Die Gegenüberstellung der beiden Skalen lässt vermuten, dass bereits bei Konzepten desselben Ursprungs, die Auffassungen des offenen Experimentierens verschieden sind. Zweifel entsteht, weil offenes Experimentieren auf der einen Skala den höchsten und auf der anderen Offenheitsskala den zweit höchsten Rang belegt. Dieser Zweifel kann mit einem Vergleich der Kategoriensets verdichtet werden. Die Gegenüberstellung führt zu dem Ergebnis, dass die Auswahl und die Anzahl der Kategorien unter dem Einfluss der Autoren stark variieren. Das hat zur Konsequenz, dass die Zuweisung eines Experimentierstils, wie offenes Experimentieren, auf unterschiedlichen Kategorien fußt. Dazu sollen einige Beispiele angeführt werden. Während Schwab (1962) und Herron (1971) eine drei- bzw. vierstufige Skala erstellen, in der beschrieben wird, ob die Kategorien ‘problem, method, answer’ (Schwab, 1962) und ‘problem area, methods of solution, interpretations’ (Herron, 1972, S. 200) offen oder vorgegeben sind, werden in den Offenheitsskalen von Hegarty (1978) die Kategorien ‘aim, materials, method, answer’ und von Hegarty-Hazel (1986 zitiert nach Staer et al., 1998) die Kategorien ‘problem, apparatus, procedure, answer’ zur Disposition gestellt. In Abgrenzung dazu werden die Öffnungsgrade bei Mayer & Ziemek (2006) aus den Kategorien ‘Fragestellung/ Hypothese, Planung, Durchführung, Auswertung, Interpretation’ und bei Buck, Bretz & Towns (2008) aus den Kategorien ‘problem/ question, theory/ background, procedures/ design, results analysis, results communication, conclusions’ abgeleitet. In Anlehnung an Mayer & Ziemek (2006) werden die Öffnungsgrade bei Hof (2010) aus den Kategorien ‘Fragestellung formulieren, Hypothese generieren, Planung des

Experiments, Durchführung des Experiments, Auswertung' generiert. Während der Experimentierstil nach dem Ansatz von Hegarty-Hazel (1986 zitiert nach Staer et al., 1998) erst dann als offenes Experimentieren klassifiziert wird, wenn alle Kategorien offen sind, wird der Experimentierstil nach dem Ansatz von Buck, Bretz & Towns (2008) schon als offenes Experimentieren klassifiziert, wenn die Kategorien 'problem/ question, theory/ background' vorgegeben werden und nur die verbleibenden Kategorien offen sind. Der Vergleich macht deutlich, dass die Auffassungen des offenen Experimentierens verschieden sind und die Ansätze von divergenten Begriffsbestimmungen begleitet werden.

Mit einem weiteren Vorgriff auf das nächste Kapitel darf berichtet werden, dass auch die Begriffsbestimmungen von Schwab (1962), Herron (1971) und deren Desideraten (Hegarty, 1978; Hegarty-Hazel, 1986; Mayer & Ziemek, 2006; Buck, Bretz & Towns, 2008 und Hof 2010 usw.) im eigenen Ansatz des offenen Experimentierens verortet werden können. Die Auswahl der Kategorien entspricht einer Auswahl von Teildimensionen, die entweder der Dimension Fachinhalt-Strategie-Methode (1. Dimension), der Dimension Phasen des Experimentierens (3. Dimension) oder beiden Dimensionen angehören. Durch die Identifikation der Kategorien als Teildimensionen und deren Transformation in den eigenen Ansatz des offenen Experimentierens wird deutlich, dass sich die Arbeiten von Domin (1999, 2007) von den bisher verorteten Begriffsbestimmungen mit festgelegten Öffnungsgraden abheben. Unterschiede bestehen, weil Domin's Differenzierung von Experimentierstilen (expository, inquiry, discovery, problem-based) nicht nur Kategorien der dritten Dimension (Phasen des Experimentierens) sondern auch Kategorien der zweiten Dimension (Lösungen und Lösungswege) enthalten. Der Autor unterteilt die Ansätze des Experimentierens danach, ob das Ergebnis vorbestimmt oder unbestimmt ist, ob der Ansatz induktiv oder deduktiv ist und ob das Vorgehen vorgegeben oder vom Lerner selbst bestimmt wird.

Ohne eine dezidierte Zuweisung aller Ansätze verfolgen zu wollen, wird stellvertretend die Begriffbestimmung von Hof (2010, S. 32) im eigenen Ansatz verortet. Die benannten Kategorien 'Fragestellung formulieren, Hypothese generieren, Planung des Experiments, Durchführung des Experiments, Auswertung' können als Teildimensionen der Experimentierphasen (3. Dimension) identifiziert werden. Kritisch muss zu dem Ansatz von Hof angemerkt werden, dass Teildimensionen wie der Fachinhalt des Experimentierens, die in den Experimentierphasen verankert sind und geöffnet werden könnten, nicht enthalten sind. Über den Ansatz von Hof hinaus muss zu allen Ansätzen des offenen Experimentierens mit festgelegten Öffnungsgraden kritisch angemerkt werden, dass ein erheblicher Dissens über die Kategorien respektive Dimensionen und Teildimensionen besteht, aus denen der Begriff des offenen Experimentierens abgeleitet wird.

Divergierende Bewertungen des offenen Experimentierens. Es folgen mehrere Bewertungen des offenen Experimentierens, denen die korrespondierenden Begriffsbestimmungen vorangestellt werden.

1. „Der Gegenpol zu detailliert angeleiteten Experimenten wäre eine offene, nicht strukturierte Aufgabenstellung.“ (Fischer & Draxler, 2001; Trendel & Fischer, 2007); „Aber auch hier zeigt sich ein geringer Wissenszuwachs. Schülerinnen und Schüler sind oft wegen fehlenden Vorwissens und geringen Möglichkeiten, das Handeln adäquat zu strukturieren, nicht in der Lage, offene experimentelle Fragestellungen physikalisch angemessen zu behandeln.“ (Fischer & Draxler, 2001, S. 390)
2. „These projects are open-ended in that the final direction and outcome are unknown at the start. The student is responsible for developing and planning the work [...]“ (Lewis, 2002, S. 139); „This study identified a number of factors, which influenced both the effectiveness of the mini-projects in preparing students for open ended project work and the extent to which students were left feeling demoralised by the experience.“ (Lewis, 2002, S. 149)
3. „In dieser Arbeit wird unter Offenheit die Möglichkeit der Schülerinnen und Schüler, eigene Entscheidungen während des Experimentierens zu treffen, verstanden.“ (Engeln, 2004, S. 40); „Das zeitlich begrenzte Erleben der völlig neuen Lernumgebung „Schülerlabor“ scheint die Schülerinnen und Schüler so sehr anzuregen und zu fordern, dass Offenheit weder auf die kognitive Aktivierung noch auf das aktuelle Interesse einen entscheidenden Einfluss hat.“ (Engeln, 2004, S. 110)
4. „Offenheit: Möglichkeit, im Rahmen von Experimentierstunden eigene Wege zu gehen.“ (Hopf, 2007, S. 89); „Selbst das Einbeziehen verschiedenster Forderungen an erfolgreiche Schülerexperimente wie Offenheit, Authentizität usw. führt immer noch nicht zu verbessertem Lernen oder positiveren Einstellungen der Schülerinnen und Schüler.“ (Hopf, 2007, S. 232)
5. „Offenheit heißt, dass in einem Unterricht mit hohem Offenheitsgrad die Rolle der Lehrkraft wechselt – von der Wissensvermittlung zur Unterstützung und Anleitung von eigenständigen Lernprozessen.“ (Hof, 2010, S. 31); „Offenes Experimentieren fördert in stärkerem Maße den Erwerb von wissenschaftsmethodischen Kompetenzen. [...] Fragend-entwickelndes Unterrichtsverfahren ist im naturwissenschaftlichen Unterricht insbesondere an solchen Stellen geeignet, an denen Problemlöse- und Experimentierprozesse angeleitet werden müssen [...]“.“ (Hof, 2010, S. 148)

Die zum Teil skeptischen Bewertungen können durch Kirschner, Sweller & Clark (2006) sowie durch Hmelo-Silver, Duncan & Chinn (2007) ergänzt werden, die die Gefahr einer Überforderung der Lernenden aufgrund der zahlreichen Anforderungen

und zu treffenden Entscheidungen als ein Kernproblem des offenen Experimentierens benennen. Im Gegensatz zu den Autoren, die die Wirksamkeit des offenen Experimentierens mit Vorbehalten sehen, kommt Reinhold (1996) auf der Basis von Fallstudien zu einer positiven Bewertung. Auch Berg, Bergendahl & Lundberg (2003) berichten, dass Experimente mit offenen Aufgaben bessere Lernergebnisse erbringen als Experimente mit geschlossenen Aufgaben. Ferner kann der Wert des offenen Experimentierens durch fachdidaktische, lern- und motivationspsychologische Argumente gestützt werden (Priemer, 2011). Beispielsweise bietet offenes Experimentieren aus der Perspektive der Lernpsychologie die Gelegenheit zu divergentem Denken. „Divergentes Denken ist ein Denken, das – entgegen den üblichen Regeln und Formen – Probleme sieht und zu lösen versucht. Das erworbene Wissen ist hierbei weniger entscheidend als die Fähigkeit, flexibel mit neuen Informationen umzugehen und zu kreativen Leistungen zu kommen“ (Schröder, 1992).

2.1.2 Eigener Ansatz des offenen Experimentierens

Die angeführten Begriffsbestimmungen und Ansätze zeigen, dass offenes Experimentieren in der fachdidaktischen Literatur bisher nicht hinreichend, sehr divergent und häufig wenig differenziert verwendet wird. Obwohl die Begriffsbestimmungen Möglichkeiten bieten, offenes Experimentieren zu differenzieren, werden die Darstellungen nicht dem Anspruch gerecht, die angeführten Dimensionen im Einzelnen detailliert zu benennen. Vielfach wird ein sehr allgemeiner Begriff zum Ausgang genommen, um die Wirksamkeit und den Wert des offenen Experimentierens zu evaluieren. Infolgedessen werden aus den Befunden Aussagen zur Überlegenheit offener bzw. nicht offener Experimentierformate generiert. Jedoch lässt ein Vergleich der Experimentieraufgaben zurzeit noch keine eindeutige Aussage zur Überlegenheit eines Aufgabenformates zu. Als eine Ursache können die unspezifischen Begriffsbestimmungen des offenen Experimentierens benannt werden, so dass unklar bleibt, auf welche Komponente des offenen Experimentierens der Erfolg bzw. Misserfolg zurückgeführt werden kann. Dazu bedarf es einer Präzisierung des Offenheitsbegriffs und der Frage, welche Dimensionen dem offenen Experimentieren zugeschrieben werden können.

Dimensionen des offenen Experimentierens. Mit der Frage, welche Dimensionen (Synonyme: Kategorien, Komponenten) dem offenen Experimentieren zugeschrieben werden können, sollen die Möglichkeiten bzw. Komponenten, die sich beim Experimentieren öffnen lassen, spezifiziert werden. Darüber hinaus ist es notwendig, dass neben den geöffneten Komponenten auch die Stärke der jeweiligen Öffnung (Synonyme: Öffnungsgrade, Graduierung) angegeben wird. Beide Aspekte fließen in die Entwicklung des eigenen Ansatzes ein, die von den zentralen Fragen

„Welche Komponenten des Experimentierens sollen geöffnet werden?“ und „Wie stark sollen die Komponenten geöffnet werden?“ geleitet wurde.

Nach dem eigenen Ansatz des offenen Experimentierens wird die Offenheit von Experimentieraufgaben in drei Dimensionen abgebildet (Priemer, Uhlmann & Kirchner, 2010; Priemer, 2011). Die drei Dimensionen sind miteinander verknüpft und zu jeder Dimension gehören mehrere Teildimensionen. Die Kriterien der Festlegung, welche Komponenten beim Experimentieren geöffnet werden sollen, lassen sich mit Hilfe von drei leitenden Fragen abbilden. Die Leitfrage der ersten Dimension lautet „Was soll geöffnet werden?“, die Leitfrage der zweiten Dimension lautet „Wie soll geöffnet werden?“ und die Leitfrage der dritten Dimension lautet „Wann soll geöffnet werden?“.

Was soll geöffnet werden? Der Fachinhalt, die Strategie und die Methode des Experimentierens werden durch die erste Dimension des offenen Experimentierens abgebildet. Der Titel dieser Dimension trägt die Bezeichnungen der Teildimensionen: Fachinhalt (1a), Strategie (1b) und Methode (1c).

Wie soll geöffnet werden? Die Anzahl der möglichen Lösungen und Lösungswege wird durch die zweite Dimension des offenen Experimentierens abgebildet. Zu dieser Dimension gehören die beiden Teildimensionen Anzahl der möglichen Lösungen (2a) und Anzahl der möglichen Lösungswege (2b).

Wann soll geöffnet werden? Die Phasen des Experimentierens werden durch die dritte Dimension des offenen Experimentierens abgebildet. Zu dieser Dimension gehören die vier Teildimensionen Aufgabenstellung (3a), Planungsphase (3b), Durchführungsphase (3c) und Auswertungsphase (3d).

Im Folgenden werden die angeführten Kategorien erklärt. Die Erklärungen geben Auskunft über die Teildimensionen und die Öffnungsstufen der Teildimensionen (Öffnungsgrade).

1. Dimension: Was soll geöffnet werden? Fachinhalt. Offenes Experimentieren kann durch die Wahl des Fachinhalts bestimmt werden. Die Öffnung des Fachinhalts ist mit der Frage verbunden, ob der Anfangszustand (die Aufgabenstellung) vorgegeben ist oder gesucht werden muss. Zwischen diesen beiden Polen der Graduierung liegt die Möglichkeit, dass aus mehreren Fachinhalten eine Auswahl getroffen werden darf. Beispielsweise kann die Bestimmung der Fallbeschleunigung vorgegeben oder in unterschiedlichen Kontexten bearbeitet werden. Eine Variation der inhaltlichen Kontexte liegt vor, wenn die Fallbeschleunigung unter Ausnutzung der Zerlegung von Kräften an der schiefen Ebene, unter Ausnutzung der gleichmäßig beschleunigten Bewegung mit der Kugelfallmethode oder unter der Ausnutzung mechanischer Schwingungen mit einem Fadenpendel bestimmt werden darf. Angemerkt sei, dass die höchste Öffnungsstufe der Teildimension – keine Vorgaben zum Fachinhalt – in der Unterrichtspraxis nur schwer zu realisieren ist, weil mit der

freien Wahl der Inhalte das Spektrum an Strategien und Methoden zur Bearbeitung des Gegenstandes zu umfangreich werden kann.

Strategie. Offenes Experimentieren ist nicht nur auf die Wahl des Fachinhalts beschränkt, sondern berührt auch die Strategien und Herangehensweisen des Experimentierens. In der Regel sollen mit der strategischen Offenheit naturwissenschaftliche Arbeitsweisen vermittelt werden, die für den Verlauf des Experimentierens bedeutsam sind. Die Offenheit kann sowohl durch qualitative als auch durch quantitative Arbeits- bzw. Experimentierweisen hergestellt werden. Folgende Beispiele seien genannt: Das Aufstellen von Hypothesen, der systematische und nicht-konfundierende Umgang mit Variablen (z. B. Variablenkontrollstrategie), die Festlegung von Verfahren zur Datenaufnahme (z. B. Anzahl der aufzunehmenden Messwerte), die Festlegung von Verfahren zum Umgang mit den Daten (z. B. Abschätzung der zu erwartenden Messunsicherheiten) und die Festlegung von Verfahren zur Datenauswertung (z. B. Vergleich nicht-konfundierender Messfolgen, Linearisierung der Daten, proportionale und antiproportionale Relationen von Messgrößen). Die Strategie des Experimentierens kann durch genaue Vorgaben, durch vorskizzierte Vorgaben und durch fehlende Vorgaben abgestuft werden. In der Regel liegen genaue Vorgaben zur Strategie vor, wenn rezeptive Versuchsanleitungen abgearbeitet werden müssen. Eine Herangehensweise könnte vorskiziert werden, indem z. B. vor Aufnahme der Experimentiertätigkeiten die Kontrolle von Einflussvariablen thematisiert wird oder Hinweise zur Aufnahme einer ausreichenden Anzahl von Messwerten gegeben werden. Wenn detaillierte oder vorskizzierte Vorgaben zur Strategie fehlen, ist die ist höchste Öffnungsstufe der Teildimension erreicht.

Methode. Damit die Experimentierstrategien verwirklicht werden können, müssen Entscheidungen zur Wahl von Messgeräten, Messanordnungen (Messvorrichtung) und Versuchsaufbauten getroffen werden. Diese Entscheidungsmöglichkeiten beim Experimentieren werden mit der methodischen Offenheit erfasst. Die Offenheit der Teildimension beschränkt sich aber nicht nur auf die Wahl der Methoden, sondern wird auch auf die Bedienung und Handhabung der Geräte übertragen. Als Messanordnungen können entweder fertige Aufbauten von Lehrmittelfirmen oder selbst entwickelte Vorrichtungen dienen, die aus den Laborgeräten von Materialsammlungen zusammengestellt werden. Die methodische Offenheit kann durch genaue Vorgaben (keine Wahlmöglichkeit der Methode), durch eine Vorselektion (Methode vorskiziert) und durch die freie Wählbarkeit (keine Vorgaben zur Methode) der Methoden unterschieden werden. Experimentieraufgaben, die sich durch eine methodische Offenheit ohne Vorgaben auszeichnen, zielen häufig auf eine Vermittlung experimenteller Fertigkeiten und Routinen. Diese kann z. B. durch einen freien Zugriff auf die Materialsammlung realisiert werden. Methodische Offenheit kann vorskiziert werden, indem aus einem bereitgestellten Pool an

Instrumenten die Messgeräte oder Messanordnungen ausgewählt werden dürfen. Hierzu seien zwei Beispiele genannt. In dem ersten Beispiel sollen verschiedene Möglichkeiten illustriert werden, wie die Drehzahl von Windrotoren für das Modell eines Auftriebsläufers bestimmt werden kann. Den angeführten Methoden liegen verschiedene Messprinzipien zugrunde. Die Drehzahl könnte mit einem Voltmeter, mit einem kommerziellen, photoelektrischen Drehzahlmesser oder mit einem Stroboskop bestimmt werden. Eine qualitative Alternative bietet die Drehzahlmessung mit Leuchtdioden. Dazu bedarf es einer elektrischen Verkettung der Leuchtdioden, so dass die Anzahl der leuchtenden Dioden mit der Messgröße variiert („Lichtleiste“). In einem zweiten Beispiel sollen mehrere Möglichkeiten zur Messung des dynamischen Auftriebs an umströmten Tragflächenprofilen (Tragflächenmodellen) präsentiert werden. Es besteht die Möglichkeit, eine Einkomponentenwaage mit Kraftmesser oder eine Digitalwaage einzusetzen, die bei Umströmung auftriebserzeugender Tragflächenprofile entlastet werden. Alternativ könnte die Digitalwaage durch eine Balkenwaage oder einen zweiseitigen Hebel ersetzt werden. Hierbei wird die Auftriebskraft gemessen, indem der Gleichgewichtszustand durch Entnahme von Massestücken wieder hergestellt wird. Wenn keine Messmethode zur Wahl gestellt wird, sondern eine Methode als verbindlich vorgegebene anzuwenden ist, dann ist die unterste Öffnungsstufe der Teildimension erreicht.

2. Dimension: Wie soll geöffnet werden? Die Offenheit einer Experimentieraufgabe lässt sich prinzipiell durch die Anzahl ihrer Lösungen (Graduierung: eine oder mehrere) und durch die Anzahl ihrer Lösungswege (Graduierung: einer oder mehrere) verankern. Hier spiegelt sich u. a. der Bezug zu Domin (1999, 2007) wider, der eine Einteilung von Experimentieraufgaben nach ihrem Ergebnis (vorbestimmt oder unbestimmt) vorgenommen hat. Mit dem eigenen Ansatz des offenen Experimentierens wird diese Klassifikation erweitert, indem die Anzahl der möglichen Lösungswege hinzugenommen wird. Die Graduierungen führen zu vier Variationen, die eine Klassifizierung offener Experimentieraufgaben zulassen. Folgende Fälle können unterschieden werden: Die Experimentieraufgabe hat eine Lösung, die auf einem Lösungsweg erreicht werden kann (Fall 1), die Experimentieraufgabe wird von mehreren Lösungen unterstützt, die auf einem Lösungsweg erreicht werden können (Fall 2), die offene Experimentieraufgabe hat eine Lösung, die auf mehreren Lösungswegen erreicht werden kann (Fall 3) und die offene Experimentieraufgabe wird von mehreren Lösungen unterstützt, die auf mehreren Lösungswegen erreicht werden können (Fall 4).

Die Fallunterscheidungen können im Sinne eines Prüfkriteriums benutzt werden, um offene Experimentieraufgaben von scheinbar offenen Experimentieraufgaben zu separieren. Mit der Begrifflichkeit „scheinbar offene Experimentieraufgaben“ werden Aufgaben gekennzeichnet, denen nur ein Lösungsweg zugrunde liegt (Fall 1

& 2). Die Offenheit dieser Experimentieraufgaben wird stark bezweifelt, weil die Adressaten bei der Lösungsfindung darauf reduziert werden, den einen existierenden Lösungsweg zu entdecken. Als ein Komplex von Experimentieraufgaben, der in der Regel dem ersten Fall folgt, seien Aufgaben mit rezeptiven Versuchsanleitungen genannt. Als ein Beispiel für den zweiten Fall sei eine Experimentieraufgabe mit gestuften Hilfen genannt (Hopf, 2007, S. 52, Aufgabenbeispiel Walkman). In dem Aufgabenbeispiel geht es thematisch um die Konstruktion einer Schmelzsicherung, die aus verschiedenen elektrischen Bauteilen (Drähte, Bimetallschalter o. ä.) gefertigt werden soll. Die Auflistung verschiedener elektrischer Bauteile zeigt an, dass zur Lösungsfindung mehrere Lösungen zugelassen sind. Aus dem Schlusssatz der letzten unterstützenden Hilfe wird allerdings ersichtlich, dass die Experimentierenden auf einem vom Autor vorgezeichneten Weg zu den Lösungen finden sollen. Die Hilfekarte enthält folgende Informationen (Hilfekarte 4 zum Aufgabenbeispiel Walkman):

„Verwende immer gleich lange und gleich geformte Drahtstücke. Verwende ein Stromstärkemessgerät, um herauszufinden, bei welcher Stromstärke der Draht durchbrennt. Die Stromstärke, für die der Walkman ausgelegt ist, kannst du aus der Spannung und der Leistung berechnen. Finde so eine geeignete Sicherung.“ (Hopf, 2007, S. 256)

Da der intendierte Lösungsweg aufgrund der Stufung des Hilfesystems für die Experimentierenden erst mit der letzten Hilfekarte ersichtlich wird, handelt es sich bei diesem Aufgabenbeispiel eher um problemorientiertes bzw. problemlösendes Arbeiten (mit einer „Scheinoffenheit“) als um eine offene Experimentieraufgabe. Als ein Beispiel für den dritten Fall kann die Materialbestimmung einer Münze benannt werden (Forschergruppe Kassel, 2006). Die Experimentieraufgabe könnte auf verschiedenen Lösungswegen wie z. B. über den statischen Auftrieb, den elektrischen Widerstand, Elektrolysen und Schmelzprozessen gelöst werden, wenn in der Aufgabenformulierung auf die zerstörungsfreie Bestimmung der Münze verzichtet werden würde. Als ein Beispiel für eine offene Experimentieraufgabe mit mehreren Lösungen und mehreren Lösungswegen (Fall 4) wird auf ein selbst entwickeltes Aufgabenbeispiel verwiesen (Kap. 2.3).

3. Dimension. Wann soll geöffnet werden? Mit der dritten Dimension wird die Offenheit einer Experimentieraufgabe in den Phasen des Experimentierens verankert. Durch die Phaseneinteilung des Experimentierprozesses in Aufgabenstellung, Planung, Durchführung und Auswertung werden die Teildimensionen festgelegt. In Übereinstimmung mit dem Fachinhalt und Prozess des Experimentierens können die Phasen durch genaue Vorgaben, durch begrenzte Vorgaben und durch fehlende Vorgaben graduiert werden. Diese Öffnungsstufen werden mit den Begriffen „Vorgehen vorgegeben“, „Vorgehen vorskizziert“ und „keine Vorgaben

zum Vorgehen“ tituliert. In der Regel liegen genaue Vorgaben zum Vorgehen vor, wenn rezeptive Versuchsanleitungen abgearbeitet werden müssen. Mit der Forderung, dass im Messprotokoll neben den Ausprägungen der Testvariablen auch die Kontrollvariablen aufgenommen werden sollen, dass zur Aufnahme von Messfolgen eine minimale Anzahl von Messdaten nicht unterschritten werden darf und dass die aufgenommenen Messfolgen nicht nur tabellarisch, sondern auch graphisch aufbereitet werden sollen, könnten z. B. die Experimentierphasen Planung, Durchführung und Auswertung vorskizziert werden. Wenn diese oder vergleichbare Vorgaben fehlen, dann ist die höchste Öffnungsstufe der Dimension erreicht.

In Anlehnung an Reinhold (1996) kann das Vorgehen auch sehr gezielt für ausgesuchte Phasen vorskizziert werden (z. B. Planungsphase), während die weiteren Teildimensionen entweder vorgegeben oder vorgabenfrei gestaltet werden dürfen. Im Bedarfsfall, der von der Wahl der offenen Experimentieraufgabe abhängt, besteht die Möglichkeit, den Experimentalprozess weiter zu partialisieren und die Offenheit einer Experimentieraufgabe in mehr als den vier genannten Teildimensionen zu verankern. Als mögliche Beispiele seien die Phasen Exploration, Hypothesen-Aufstellen, Aufbau von Experimenten, Interpretation und Präsentation der Ergebnisse genannt.

Der hier vorgestellte Ansatz trägt somit zu einer präziseren Fassung und Begriffsbildung des offenen Experimentierens bei. Die Präzisierung der Begriffsbildung gelingt, weil explizit die Öffnungsmöglichkeiten (Dimensionen) und die Öffnungsstärken (Graduierung) von Experimentieraufgaben benannt werden. Die Offenheit der Lernumgebung „Experiment“ wird durch drei Dimensionen und mehrere, mehrfach gradierte Teildimensionen abgebildet. Die Öffnungsstärken sind den Teildimensionen angepasst worden, so dass ihre Ausprägungen von Teildimension zu Teildimension variieren. Die Entwicklung eines eigenen Ansatzes wurde notwendig, weil der Begriff des offenen Experimentierens in der fachdidaktischen Literatur zu allgemein, zu unspezifisch und nicht erschöpfend gefasst wird. Den Problemen wird mit einer angemessenen Begriffsbildung begegnet, indem die Offenheit von Experimentieraufgaben nicht nur präzise beschrieben sondern auch maßvoll dimensioniert wird. Eine überdimensionierte und nicht angemessene Begriffsbestimmung läge vor, wenn die Kategorisierungen umfassender gefasst würden als es zur Beschreibung offener Experimentieraufgaben notwendig wäre. Beispielsweise ist die Frage, ob der Lehrer oder der Lerner (Schüler) für die Offenheit beim Experimentieren sorgt (organisatorische Offenheit) für die Begriffsbildung nicht relevant, da Entscheidungen zur organisatorischen Offenheit den Ansätzen zum offenen Unterricht und nicht den Ansätzen zum offenen Experimentieren zuzurechnen sind.

1. Dimension: Fachinhalt-Strategie-Methode (Was soll geöffnet werden?)

Teildimension	Graduierung
1a) Gegenstand	Fachinhalt vorgegeben
	mehrere Fachinhalte zur Wahl
	keine Vorgaben zum Fachinhalt
1b) Strategie	Strategie vorgegeben
	Strategie vorskizziert
	keine Vorgaben zur Strategie
1c) Methode	Methode vorgegeben
	Methode vorskizziert
	Keine Vorgaben zur Methode

2. Dimension: Lösungen und Lösungswege (Wie soll geöffnet werden?)

Teildimension	Graduierung
2a) Lösungen	eine Lösung
	mehrere Lösungen
2b) Lösungswege	ein Lösungsweg
	mehrere Lösungswege

3. Dimension: Phasen des Experimentierens (Wann soll geöffnet werden?)

Teildimension	Graduierung
3a) Aufgabenstellung	Vorgehen vorgegeben oder Vorgehen vorskizziert oder keine Vorgaben zum Vorgehen
3b) Planung	
3c) Durchführung	
3d) Auswertung	

Abb. 1: Die Dimensionierung und Graduierung des offenen Experimentierens nach dem eigenen Ansatz im Überblick

In Abgrenzung zu den Begriffsbestimmungen von Hof (2010), Mayer & Ziemek (2006), Millar, Tiberghien & Le Maréchal (2002), Domin (1999), Hegarty-Hazel (1986) und anderen ist deshalb die organisatorische Offenheit im eigenen Ansatz nicht verankert worden. Wie am Beispiel von Hof (2010), Mayer & Ziemek (2006),

Millar, Tiberghien & Le Maréchal (2002), Domin (1999), Hegarty-Hazel (1986) u.a. aufgezeigt werden kann, gelingt es dennoch auch überdimensionierte Begriffsbestimmungen des offenen Experimentierens dem eigenen Ansatz zuzuordnen. Der eigene Ansatz wird mit der Abbildung 1 zusammengefasst.

2.1.3 Perspektiven des offenen Experimentierens

Der vorausgehende Ansatz des offenen Experimentierens bietet die Möglichkeit, Experimentieraufgaben maßvoller konstruieren und analysieren zu können. Dieser didaktische Mehrwert entsteht, weil explizit die Öffnungsmöglichkeiten (Dimensionen) und Öffnungsstärken (Graduierung) von Experimentieraufgaben benannt werden, die eine präzisere Konstruktion, Typisierung und Beschreibung von Experimentieraufgaben erlauben. Infolgedessen werden zur Klassifizierung experimenteller Aufgabenformate begriffserweiternde Attribute wie selbstständig (Reinhold, 1996), nicht strukturiert (Fischer & Draxler, 2001; Trendel & Fischer, 2007) und problemorientiert (Mikelskis-Seifert & Rabe, 2007) hinfällig. Während Priemer (2011) für mehrere publizierte Aufgabenbeispiele aufzeigt, dass Experimentieraufgaben nach dem eigenen Ansatz präziser typisiert werden können, wird in der vorliegenden Arbeit mit einem selbst entwickelten Aufgabenbeispiel aufgezeigt (Kap. 2.3), dass Experimentieraufgaben nach dem eigenen Ansatz des offenen Experimentierens präziser konstruiert werden können. Der Ansatz trägt dazu bei, dass passgenaue Designs offener und geschlossener Experimentieraufgaben erstellt werden können, die mit den bisher bekannten Konzepten des offenen Experimentierens nicht hinreichend beschrieben werden können. Es darf daher propagiert werden, dass Experimentieraufgaben nicht mehr über deren graduelle Ausprägungen wie offen, partiell offen oder formal (Kirschner, 1992 zitiert nach Engeln, 2004), sondern über deren Dimensionierung und Graduierung definiert werden müssen. Der Ansatz bietet zudem die Chance, prozessbegleitende Hilfen gezielter und Hilfesysteme adressatengerechter formulieren zu können. Beispielsweise könnten die Lernenden mit Unterstützung von schriftlichen Hilfen dazu befähigt werden, die Art der benötigten Hilfen aus einem umfassenden Hilfeangebot auszuwählen. Dieses Ziel setzt allerdings voraus, dass die offerierten Hilfen mit den angegebenen Dimensionen des Experimentierens (vgl. eigener Ansatz) abgestimmt werden. Die Maßnahme führt zu einem Hilfesystem, welches zwischen inhaltlichen, strategischen und methodischen Hilfearten differenziert. Zudem könnten mit den Hilfen auch mögliche Lösungen und Lösungswege oder ausgesuchte Experimentierphasen (z. B. Planung oder Auswertung) fokussiert werden. Ein entscheidender Aspekt hierbei ist, dass den Lernenden die verschiedenen Hilfearten bekannt sein und ihnen ein direkter Zugriff gewährt werden muss.

Darauf aufbauend kann die Wirksamkeit einzelner und das Zusammenspiel mehrerer Dimensionen untersucht werden, denn die bekannten Befunde lassen zurzeit noch keine eindeutigen Aussagen zur Überlegenheit offener bzw. geschlossener Aufgabenformate zu. Was eine Beurteilung der Studien erschwert, ist die Verwendung zum Teil sehr unterschiedlicher Lernsettings und Maßstäbe für die Bewertung des Experimentiererfolgs. Beispielsweise sind verschiedene Ansätze bekannt mit denen in experimentellen Settings eine Überforderung der Lernenden vermieden werden soll, ohne sie dadurch zu unterfordern (z. B. Chen & Klahr, 1999; Klahr & Nigam, 2004; Hammann, Phan & Bayrhuber, 2007; Kap. 3 ff. in dieser Arbeit). Viele davon nutzen schriftliche Anleitungen und andere geleitete Hilfen (Forschergruppe Kassel, 2006, 2007; Hopf, 2007; Wieczorek & Sommer, 2008). Die Ursachen der divergierenden Befunde zu offenen und geschlossenen Experimentieraufgaben liegen aber nicht nur in den unspezifischen Begriffsbestimmungen, sondern auch darin, dass die Art der Experimentieraufgaben (offen versus geschlossen) und die zu vermittelnden Lerninhalte wie inhaltliches, strategisches und methodisches Wissen, nicht adäquat aufeinander abgestimmt sind. Dieser Aspekt berührt auch die Frage, welche Lernziele erreicht und welche Kompetenzen bei den Schülern aufgebaut werden sollen. Lunetta, Hofstein & Clough (2007) fassen als „Principal Goals for Learning in the School Laboratory“ zusammen:

„Promote the development of students’ scientific knowledge, problem-solving abilities, and habitus of mind, including: (1) Conceptual knowledge, (2) Practical skills and problem-solving abilities; now expanded to include: argumentation from data (*procedural knowledge*), (3) Knowledge of how science and scientists work, (4) Interest and motivation, (5) Understanding methods of scientific inquiry and reasoning; now expanded to include the nature of science.” (Lunetta, Hofstein & Clough, 2007, S. 402)

Die Ziele von Lunetta et al. (2007) sind eher allgemein für den naturwissenschaftlichen Unterricht formuliert worden. Genauere Aufstellungen sind beispielsweise bei Klopfer (1990), Millar (1998) sowie Singer, Hilton & Schweingruber (2006) zu finden. Ebenfalls spezifischer auf das Experimentieren ausgerichtet als bei Lunetta et al. (2007) beschreiben die Bildungsstandards für den Mittleren Schulabschluss die zu erwerbenden Kompetenzen (KMK 2005a, 2005b, 2005c). In den naturwissenschaftlichen Fächern sind im Kompetenzbereich Erkenntnisgewinnung einheitlich und explizit die Nutzung experimenteller Untersuchungsmethoden vorgesehen. Beispielsweise wird in den Bildungsstandards Physik angeführt: „Eingebettet in den Prozess physikalischer Erkenntnisgewinnung sind das Experimentieren und das Entwickeln von Fragestellungen wesentliche Bestandteile physikalischen Arbeitens“ (KMK 2005c, S. 10). Die biologische-, chemie- und physikdidaktische Forschung hat zur systematischen Beschreibung von Experimentierkompetenzen entsprechende Messverfahren entwickelt. Es liegen

damit Kompetenzmodelle vor, die sich zum Teil noch in der Validierungsphase befinden (vgl. Schecker, Fischer & Wiesner, 2004; Schecker & Parchmann, 2006; Hammann, Ganser & Haupt, 2007; Mayer, Grube & Möller, 2008; Klos, Henke, Kieren, Walpuski & Sumfleth, 2008; Walpuski, Kampa, Kauertz & Wellnitz, 2008; Labudde, Metzger & Gut, 2009; Mannel, Sumfleth & Walpuski, 2009; Nawrath, Maisyenko & Schecker, 2011).

Die angeführten Arbeiten legen nahe, die Lernziele in dreierlei Hinsicht zu differenzieren: Inhaltlich – die Aneignung fachlichen Wissens und Verstehens, strategisch – die Aneignung von strukturierten und systematischen Vorgehensweisen beim Experimentieren, fachmethodisch – die Aneignung von Methoden und Routinen des Experimentierens. Die Betrachtung der durch Experimente zu vermittelnden Lernziele und Kompetenzen zeigt, dass diese deutlich über rein fachmethodische Fertigkeiten hinausgehen, insbesondere die Struktur und das Vorgehen naturwissenschaftlichen Arbeitens in den Vordergrund rücken (strategische Aspekte) und nicht zuletzt auch den inhaltlich-konzeptuellen Wissenserwerb im Blick haben. Damit wirft die Mehrzahl an zu vermittelnden Kompetenzen die Frage nach der jeweils geeigneten didaktischen Methode auf (offene versus geschlossene Experimentieraufgaben). Hieraus ergibt sich eine weitere Perspektive für den eigenen Ansatz des offenen Experimentierens.

2.2 Naturwissenschaftliche Arbeitsweisen

Ob Kenntnisse in den Naturwissenschaften Bildung ausmachen, war eine Frage, die vor mehr als hundert Jahren noch diskutiert wurde. Als eine immanente Antwort auf die historisch bedeutende Frage dürfen die epochal gefärbten, fachwissenschaftlich, pädagogisch und didaktisch beeinflussten Ansätze dessen, was man von den Naturwissenschaften wissen sollte (Art, Umfang, Vermittlungsziele), gewertet werden. Beispielsweise sei die lange Zeit vertretene Maxime des naturwissenschaftlichen Unterrichts genannt, die Methode des Unterrichts an der Methode der Wissenschaft zu orientieren (Duit, Gropengießer & Stäudel, 2007), was in Mothes „Normalverfahren“ zum Ablauf von Unterricht mündete (Mothes, 1968) und zu einem aus heutiger Perspektive widerlegten, einheitlichen Grundschema naturwissenschaftlicher Erkenntnisgewinnung führte.

Eine neue Epoche in der Bildungsdiskussion wurde mit dem Modell der Scientific Literacy (z. B. Bybee, 1997) und dem Beitrag der Naturwissenschaften zur allgemeinen Bildung (z. B. Gräber, Nentwig, Koballa & Evans, 2002) angestoßen. Ein Kennzeichen des Wandels ist die Akzentuierung der naturwissenschaftlichen Arbeitsweisen, die, anders als bisher, zum Gegenstand naturwissenschaftlicher

Grundbildung gemacht werden, ohne die Vermittlung naturwissenschaftlicher Begriffe und Prinzipien zu vernachlässigen. Diese Dualität vom Wissen über die Naturwissenschaften und Wissen in den Naturwissenschaften wird z. B. mit dem vierstufigen Modell einer naturwissenschaftlichen Grundbildung von Bybee (ebd.), dem Kompetenzmodell von PISA (Prenzel, Artelt, Baumert et al., 2007), den nationalen Bildungsstandards für den Mittleren Schulabschluss (KMK 2005a – c) und den landesspezifischen Kern- und Rahmenplänen (z. B. Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008) transportiert.

Obwohl die Popularität der Arbeiten von Bybee (ebd.) durch ihren Einfluss auf das Kompetenzmodell von PISA gesteigert werden konnte (vgl. Schecker & Parchmann, 2006), wurde in den Didaktiken der Naturwissenschaften bereits lange vor Beginn des PISA-Zeitalters zwischen Konzepten, die für die wissenschaftlichen Begriffe und Prinzipien stehen und Prozessen, mit denen die naturwissenschaftlichen Denk- und Arbeitsweisen bezeichnet werden, unterschieden (Duit, 2003). Eine umfassende Übersicht zur Unterscheidung von Konzept- und Prozesszielen, die mit dem Experiment zu tun haben, wurde von Klopfer (1971) erstellt.

2.2.1 Prozessbezogene Fähigkeiten des Experimentierens

Modellierung. Im Zuge der Diskussion um Scientific Literacy hat die Modellierung von Erkenntnisprozessen beim Experimentieren an Bedeutung gewonnen, was in jüngster Vergangenheit zu einer Anhäufung von Kompetenzmodellen geführt hat (vgl. Kap. 2.1.3). Bemerkenswert ist, dass die Beschreibung der Erkenntnisgewinnung beim Experimentieren auch ein Anliegen der Kognitionspsychologie ist und deren Modellbildungen von der biologische-, chemie- und physikdidaktischen Forschung wohlwollend adaptiert werden, obwohl die Ausrichtung der Forschungsdisziplinen anderen Interessen folgen dürften. In diesem Zusammenhang sei das momentan sehr populäre SDDS-Modell genannt (Klahr & Dunbar, 1988; Klahr, 2000), das den naturwissenschaftlichen Erkenntnisprozess aus der Sicht der Problemlöseforschung abbildet. Mit dem Akronym SDDS sollen zwei Räume, der Hypothesenraum und der Experimentierraum, akzentuiert werden, in denen nach Problemlösungen gesucht werden muss. Es steht für Scientific Discovery as Dual Search.

Die Hauptkomponenten des SDDS-Modells (top level) sind die „Suche im Hypothesenraum“, die „Suche im Experimentierraum“ und die „Analyse von Evidenzen“. Der Namensgebung folgend wird der Prozess der Hypothesenbildung von der „Suche im Hypothesenraum“ erfasst, während der sich anschließende Teilaspekt des Testens von Hypothesen der „Suche im Experimentierraum“

vorbehalten ist. Output dieser Komponente sind die Evidenzen, die zur Bewertung der aufgestellten Hypothesen notwendig sind, während die Entscheidungen über die Akzeptanz, über den Widerruf oder eine weitere Prüfung der aufgestellten Hypothesen im Bereich der „Analyse von Evidenzen“ liegen. Dieser Komponente wird eine vermittelnde Funktion zwischen beiden Räumen zugeschrieben, von der aus die weitere Suche im Hypothesen- bzw. Experimentierraum gesteuert wird. Jenseits der top level Komponenten des SDDS-Modells existiert eine sehr detaillierte Unterstruktur, die bei Klahr (2000) eingehend beschrieben wird und aufgrund der aktuellen Präsenz des SDDS-Modells in der Literatur (z. B. Walpuski, 2006; Hammann, 2007; Schreiber, Theyßen & Schecker, 2009; Emden, 2011) nicht mehr näher erläutert wird.

Ein Indikator dafür, dass der Prozess der Hypothesenbildung mit dem SDDS-Modell noch nicht hinreichend systematisiert worden ist, wird durch die erweiterte Fassung des SDDS-Modells angezeigt. Sie wurde von de Jong und van Joolingen (van Joolingen & de Jong, 1997; de Jong & van Joolingen, 1998) ausgearbeitet und als Extended SDDS-Modell publiziert. Nach dem Extended SDDS-Modell ist im SDDS-Modell das Verhältnis von Schülerhypothesenraum, Schülerwissensraum und Lernzielraum nicht adäquat berücksichtigt worden. Diese Korrektur des SDDS-Modells schafft eine Voraussetzung, damit auch der wissenschaftliche Erkenntnisprozess einer kollaborativ forschenden Gruppe beschrieben werden kann, bei der die Schnittmenge der individuellen Wissensräume in den Blickpunkt genommen werden muss (Okada & Simon, 1997).

Diagnoseinstrumente. Zur Messung und Beschreibung prozessbezogener Fähigkeiten beim Experimentieren werden von den Didaktiken der Naturwissenschaften und der Lehr-Lern-Forschung sehr unterschiedliche Messinstrumente bereitgehalten. Das Spektrum an Möglichkeiten erstreckt sich von der direkten Beobachtung des Experimentierens bis zum Multiple-Choice-Test. Ohne eine Abstufung der Erhebungsmethoden vornehmen zu wollen, können zwischen diesen beiden Varianten Testformate wie Computersimulationen, Laborjournale, Protokolle und Kurzantwortentests u. a. eingruppiert werden. Eine sinnvolle Einteilung ergibt sich, wenn zwischen paper-pencil-basierten Testformaten, Experimentiertests mit Computersimulationen und Experimentiertests mit Realexperimenten (performance assessment, vgl. Ruiz-Primo & Shavelson, 1996; Stebler, Reusser & Ramseier, 1998) unterschieden wird, denen im Folgenden eine exemplarische Auswahl von Erhebungen zugeordnet wird.

Als Beispiele für Erhebungen, die schriftliche Testformate eingesetzt haben, können Large-Scale-Studien wie TIMSS (Baumert, Lehmann, Lehrke et al., 1998; Baumert, Bos, Klieme et al., 1999) oder PISA (Prenzel, Artelt, Baumert et al., 2007) benannt werden, während Experimentiertests mit Computersimulationen von Baxter & Shavelson (1994), Shavelson, Ruiz-Primo & Wiley (1999) oder Schreiber, Theyßen

& Schecker (2009, 2011) benutzt wurden und Experimentiertests mit Real-experimenten im Rahmen des TIMSS Performance Assessments (Harmon et al., 1997; Stebler, Reusser & Ramseier, 1998; Labudde & Stebler, 1999) oder im schweizerischen Projekt HarmoS bevorzugt wurden (Zeyer, Adamina, Gingins & Labudde, 2007; Wissenschaftliches Konsortium HarmoS Naturwissenschaften+ 2008).

Im deutschsprachigen Raum wurde die Diskussion um die Konstruktion und den Einsatz von Experimentiertests mit der Forderung der Klieme-Expertise zur Entwicklung von Bildungsstandards belebt (Klieme, Avenarius, Blum, Döbrich, Gruper, Prenzel et al., 2003), nach der die aufgestellten Kompetenzen grundsätzlich überprüfbar sein sollen. Da die Umsetzung der Forderung für den Kompetenzbereich Erkenntnisgewinnung eine Herausforderung darstellt, wird in den Didaktiken der naturwissenschaftlichen Disziplinen die Austauschbarkeit von Testformaten diskutiert. Eine zentrale Frage, die sich aus der Messung experimenteller Fähigkeiten ergibt, ist, welche prozessbezogenen Kompetenzen des Experimentierens von den vorgestellten Diagnoseinstrumenten (s. o.) erfasst werden können, so dass sie als Ersatzmaß für die direkte Beobachtung nutzbar gemacht werden können. Die Relevanz der Frage ergibt sich aus dem Forschungsanliegen, die Auswertung prozessbezogener Fähigkeiten des Experimentierens ökonomischer zu gestalten als es die direkte Beobachtung momentan zulässt. Um den Grad des erfüllten Forschungswunsches wissenschaftlich zu fundieren, werden vielfach Korrelationsanalysen erstellt, mit denen die Passung zwischen der direkten Beobachtung als Referenzmaß und dem substituierenden Testformat bestimmt wird (z. B. Baxter & Shavelson, 1994; Shavelson, Ruiz-Primo & Wiley, 1999; Emden, 2011). Unter Auslassung der erzielten statistischen Übereinstimmungen können grobstufige Kategoriensysteme, produktbezogene Datenauswertungen oder divergierende Testzeitpunkte identifiziert werden (vgl. Schreiber, Theyßen & Schecker, 2009), die die Aussagekraft der Korrelationsanalysen dämpfen.

Als eine kompensatorische Maßnahme mit hoher Akzeptanz könnte sich die Einführung von Zielvereinbarungen erweisen, mit denen prozessbezogene Fähigkeiten des Experimentierens als Austauschbedingung für Testformate festgeschrieben werden. Zu diesem Zweck könnten beispielsweise die lernwirksamen Merkmale zur Beschreibung von Unterrichtsqualität auf das Experimentieren übertragen werden (Reyer, Trendel & Fischer, 2004; Trendel, Wackermann & Fischer, 2007). Den Befunden der Lehr-Lern-Forschung folgend (Pauli & Reusser, 2003; Klieme, Lipowsky, Rakoczy & Ratzka, 2006), dürften dann die Sichtstrukturen des Experimentierens eher geringere Auswirkungen auf die Lernergebnisse haben, während die Tiefenstrukturen des Experimentierens und deren Wechselwirkung mit der Sichtstruktur als obligatorische Elemente in den Austauschbedingungen für Testformate festgehalten und von den aufgeführten

Diagnoseinstrumenten (s. o.) erfasst werden müssten. Aller Voraussicht nach können diese Erwartungen von schriftlichen Testformaten wie Paper-Pencil-Tests nicht erfüllt werden, von denen bereits bei geringeren Ansprüchen bekannt ist, dass sie die prozessbezogenen Fähigkeiten des Experimentierens nur unzureichend abbilden und mit den Ergebnissen direkter Beobachtungen eher schwach korrelieren (vgl. Shavelson, Ruiz-Primo & Wiley, 1999; Hammann, Phan & Bayrhuber, 2007).

Als eine Alternative, mit der die Verwendung von Realexperimenten aufrecht erhalten werden würde, wird in der fachdidaktischen Literatur der Einsatz einer Protokollmethode für die prozessorientierte Leistungsmessung diskutiert (z. B. Emden, 2011; Emden & Sumfleth, 2012). Hierbei darf jedoch nicht vergessen werden, dass die Erhebungsmethode auf einer schriftlichen Selbstauskunft der Probanden basiert, die jenseits der Wiedergabe von Messdaten beim Experimentieren kaum Erfahrungen im Umgang mit dem Diagnoseinstrument haben. Ein vorbereitendes Training zum Gebrauch des Messinstruments scheint daher unumgänglich, was nach den Erfahrungen von Emden (ebd.) auch vorgeschlagen wird. Ein ungelöstes Problem, was die Verwendungsmöglichkeiten der Datenerhebungsmethode begrenzt, ist, dass ihr Einsatz nicht vom Training zum Gebrauch des Messinstruments entkoppelt werden kann. Beispielsweise können die Probanden keine schriftlichen Selbstauskünfte über die Planung, Durchführung und Auswertung konfundierender Experimente erteilen (protokollieren), wenn ihnen Kenntnisse über die Variablenkontrollstrategie (s. u.) fehlen. Andererseits werden die Probanden mit einem vorbereitenden Training auf diese und andere defizitäre Verhaltensweisen beim Experimentieren aufmerksam gemacht, was zu erheblichen Verzerrungen bei der Datenerhebung führen würde. Daraus ergibt sich die kritische Frage, welche der Probleme und Schwierigkeiten für den Umgang mit Variablen beim Experimentieren (vgl. Kap. 2.2.2) nach einem vorbereitenden Training mit der Protokollmethode noch zuverlässig gemessen werden könnten.

Aufgrund dieser und weiterer Probleme mit denen die Verwendung von Ersatzmaßen behaftet ist, werden die prozessbezogenen Fähigkeiten beim Experimentieren in der vorliegenden Studie von trainierten Datenaufnehmern unter Einsatz eines selbst erstellten Messinstruments (Beobachtungsbogen, vgl. Kap. 4.2.3) direkt beobachtet. Eine begrenzte Auswahl von Probanden wird zusätzlich videografiert.

2.2.2 Der Umgang mit Variablen beim Experimentieren

Der Umgang mit Variablen beim Experimentieren ist eine wichtige Arbeitsweise der naturwissenschaftlichen und naturwissenschaftsverwandten Disziplinen, die für den Biologie-, Chemie- und Physikunterricht von grundlegender Bedeutung ist. Trotz der domänenübergreifenden Eigenschaften, kann die Ausprägung der Arbeitsweise

fachspezifische Unterschiede annehmen. Von den naturwissenschaftlichen Disziplinen darf für die Physik konstatiert werden, dass aus vielen Teilbereichen umfassende Aspekte der Arbeitsweise nutzbar gemacht werden können. Im Kontext von Schule liegt die Relevanz der fachspezifischen Methode in ihrer experimentierstilübergreifenden Verwendung, die in den wesentlichen Prozessschritten des Experimentierens (Planung, Durchführung, Auswertung) verankert ist und als prozessbezogene Fähigkeit in mehreren Kompetenzmodellen abgebildet wird. Als ein Beispiel sei das in der Kognitionspsychologie verankerte SDDS-Modell genannt, das dem Umgang mit Variablen bereits in den top level Komponenten Suche im Hypothesenraum, Suche im Experimentierraum und Analyse von Evidenz einen hohen Stellenwert beimisst (vgl. Kap. 2.2.1). Obwohl die Kompetenzausprägungen für den Umgang mit Variablen von Modell zu Modell variieren, können gemeinsame, modellübergreifende Schlüsselqualifikationen bestimmt werden. Unter den Experimentierstrategien wird vor allem die Variablenkontrollstrategie als eine Kernkompetenz benannt, deren Untersuchungsinteressen über die fachdidaktische Literatur hinausgehen (s. o.).

Aus fachdidaktischer Perspektive sind besonders jene Forschungsergebnisse wertvoll, die die Probleme und Defizite der Experimentierenden beim Umgang mit Variablen beschreiben und einen direkten oder indirekten Bezug zu den Prozessschritten des Experimentierens herstellen. Als Kernproblem erweist sich eine Experimentierweise, bei der die Lernenden die Wirkung von Testvariablen untersuchen, ohne die Kontrollvariablen konstant zu halten. Dieser Verstoß gegen die Variablenkontrollstrategie, der als konfundierende Experimentierweise bezeichnet wird, zeigt an, dass zwischen den beiden Variablentypen nicht unterschieden wird und ggf. auch Probleme mit der Isolierung von Variablen vorherrschend sind (Klayman & Ha, 1989). Im Extremfall werden alle Variablen gleichzeitig verändert. Für diese Experimentierstrategie wurde die Bezeichnung „change all“ vergeben (Tschirgi, 1980). Ein weiteres Defizit, dass beim Planen von Experimenten identifiziert werden kann, ist ein Fehlen von aussagekräftigen und die Wahl von falschen Kontrollansätzen. Beide Mängel erweisen sich als hinderlich, wenn Messfolgen angelegt werden. Hier kann festgestellt werden, dass die Ausprägungen von Variablen unsystematisch verändert und Experimentieransätze unlogisch verknüpft werden (Kuhn & Angelev, 1976; Schauble, Klopfer & Raghavan, 1991). In diesem Zusammenhang muss auch eine Vorgehensweise erwähnt werden, die häufig mit einer unsystematischen Verwendung von Variablen einhergeht und als „local chaining“ bezeichnet wird. Darunter zu verstehen ist ein isolierender Vergleich von direkt aufeinanderfolgenden Messwertpaaren, die verschiedenen, mal konfundierenden, mal nicht-konfundierenden Messfolgen angehören und nicht aufeinander bezogen werden (Schauble, 1996). In der Summe weisen diese Befunde darauf hin, dass Experimentierende weniger dazu neigen, die

Wirkung der beteiligten Variablen zu ergründen, als bestimmte Effekte zu erzeugen (Bullock & Ziegler, 1999; Sodian, Bullock & Koerber, 2008). Zu dieser Einschätzung kommt auch Hammann (2004), die er mit Rückbezug auf Carey, Evans, Honda, Jay & Unger (1989) und Schauble, Klopfer & Raghavan (1991) in einer sehr generalisierenden Aussage zusammenfasst: „Sie sollen unter anderem lernen, wie Naturwissenschaftler und nicht wie Ingenieure zu experimentieren“ (Hammann, Phan, Ehmer & Bayrhuber, 2006, S. 292). Hinsichtlich der Analyse und Auswertung von Messdaten ist ebenfalls bekannt, dass Lernende dazu tendieren, lediglich bestätigende Daten zu erheben und Evidenzen, die den Erwartungen widersprechen, ignoriert oder umgedeutet werden (Chinn & Brewer, 1998). In der Literatur wird dieser Effekt mit dem Begriff „confirmation bias“ gekennzeichnet (Wason, 1960; Klayman & Ha, 1989). Für detailliertere Angaben, die fehlerbehaftete oder defizitäre Experimentierweisen beschreiben, sei auf eine Übersicht von Hammann (2004) und Hammann, Phan, Ehmer & Bayrhuber (2006) verwiesen, die mit vielen Beispielen unterlegt worden ist.

Aufgrund der zahlreichen Probleme und Defizite, die beim Umgang mit Variablen registriert worden sind und direkt oder indirekt auf Verstöße gegen die Variablenkontrollstrategie zurückgeführt werden können, muss hinterfragt werden, ob bei den Lernenden die psychologischen Voraussetzungen zum Einsatz der Strategie hinreichend entwickelt sind. Mit Bezug auf Lawson & Wollmann (1976) sowie auf Chen & Klahr (1999) schreibt Hammann (2004, S. 199): „7–10 – Jährige können durch explizites Training die systematische Kontrolle von Variablen bereits erlernen und unkonfundierte Experimente in der gleichen Domäne erfolgreich anlegen; 10-Jährige übertragen die erworbenen Strategien bei der Planung von Experimenten auf andere Domänen.“ Eine differenzierte Analyse, die die Entwicklung der Denk- und Argumentationsfähigkeiten von der Kindheit bis zum Erwachsenenalter hinterfragt und sich insbesondere die Entwicklung der Variablenkontrollstrategie als Schlüsselkompetenz des systematischen Experimentierens zu eigen gemacht hat, wurde im Rahmen der Münchener Längsschnittstudie LOGIK erstellt (Sodian, Bullock & Koerber, 2008). Die wesentlichen Befunde der Studie waren, dass die meisten Kinder im Grundschulalter spontan noch nicht die Variablenkontrollstrategie zeigten (ebd., S. 73), nur 36 % der 12-Jährigen (6. Klasse) aber 76 % der 18-Jährigen spontan gute Experimente konstruierten (ebd., S. 76), indem sie die fokale Dimension variierten und alle anderen Dimensionen konstant hielten. Wesentlich bessere Ergebnisse konnten mit einer Variante des eingesetzten Aufgabenbeispiels erzielt werden. Wenn den Probanden die Variationsmöglichkeiten eines Gegenstandes aufgezeigt wurden und sie anschließend gebeten wurden, ein kontrolliertes Experiment zu wählen, dann wurde die Aufgabe von 40 % der Drittklässler, 65 % der Viertklässler, annähernd 70 % der Fünftklässler, über 80 % der Sechstklässler, über 90 % der 18-Jährigen und etwa 95 % der 23-Jährigen erfüllt

(ebd., S. 75). Auf Basis dieser und weiterer Ergebnisse (Zimmerman, 2007) kommen die Autoren zu dem Schluss, dass die Fähigkeit zum wissenschaftlichen Denken nicht erst in der Adoleszenz entsteht. Zudem stützen sie ihren Befund mit einem Effekt, den ein Strategietraining auf die Produktion der Variablenkontrollstrategie bei Drittklässlern hatte (Bullock & Ziegler, 1999).

Als fachdidaktischer Ertrag ergibt sich aus den entwicklungspsychologischen Untersuchungen, dass die defizitären Verhaltensweisen beim Umgang mit Variablen und Verstöße gegen die Variablenkontrollstrategie durch Unterstützungsmaßnahmen kompensiert werden können. Allerdings sind die Studien nicht mit Realexperimenten durchgeführt worden, so dass die Kernfrage „how much and what kind of guidance to provide?“ (Mayer, 2004, S. 17) nach wie vor unbeantwortet bleibt. Hier liegt die Aufgabe in der Entwicklung und Balancierung von alters- bzw. experimentiertypenspezifischen Instruktionen, die eine Überforderung der Lernenden vermeiden, ohne sie dadurch zu unterfordern. Aus der fachdidaktischen und psychologischen Literatur sind zahlreiche Unterstützungsmaßnahmen bekannt, die für den Umgang mit Variablen beim Experimentieren erprobt werden könnten. Als Beispiele seien gestufte Lernhilfen (z. B. Stäudel, Franke-Braun & Schmidt-Weigand, 2007), die Darbietung von Prompts (z. B. Lin & Lehman, 1999; Thillmann, Küsting, Wirth & Leutner, 2009), strategische Trainings (z. B. Klahr & Nigam, 2004), Strukturierungshilfen (z. B. Walpuski, 2006), Fehlerkorrektur (z. B. Walpuski, 2006), die Kombination aus Strukturierungshilfen und Lehrervortrag (z.B. Wahser & Sumfleth, 2008) und wissenschaftstheoretischer Unterricht (z. B. Sodian, Jonen, Thoermer & Kircher, 2006) genannt, die je nach Ansatz, Einsatz und Verwendung in explizite und implizite Instruktionsmaßnahmen eingeteilt werden können.

Abschließend muss für den Umgang mit Variablen beim Experimentieren kritisch angemerkt werden, dass in der Literatur die evidenzbasierte Datenauswertung vergleichsweise wenig untersucht worden ist. Es darf bemängelt werden, dass vor allem solche Aufgabenbeispiele zum Gegenstand fachdidaktischer und psychologischer Untersuchungen gemacht werden, die entgegen anderslautender Forderungen (Kanari & Millar, 2004) eine Konfrontation der Experimentierenden mit verschiedenen Variablenverläufen (linear, extremal, ohne Einfluss) und unterschiedlichen Effektstärken vermeiden. Dem Schwerpunkt nach müssen in wissenschaftlichen Untersuchungen Messdaten ausgewertet werden, die die Probanden mit Hilfe präparierter (vorgefertigter) Experimentiermaterialien am Beispiel von Variablen mit linearem Verlauf generiert haben. Bei den vorgefertigten Versuchsmaterialien handelt es sich häufig um eine Auswahl von Gegenständen wie Federn (Chen & Klahr, 1999), Auftriebskörper (Küsting, Thillmann, Wirth, Fischer & Leutner, 2008), schiefe Ebenen (Baumert et al., 1998, Item I 12; Klahr & Nigam, 2004), die in realen Experimentierumgebungen, in computerbasierten Lernsettings

oder in schriftlichen Aufgabenformaten zu nicht-konfundierenden Messfolgen kombiniert werden sollen. Üblicherweise werden aus dem (Experimentier-)verhalten der Probanden die Konzepte und Strategien zum Umgang mit Variablen ausgelesen und als wissenschaftlich erhärtete bzw. konsolidierte Befunde eingestuft, wenn sie in verschiedenen (Experimentier-)Kontexten gemessen werden können (z. B. Schauble, Klopfer & Raghavan, 1991, Bullock & Ziegler, 1993; Chen & Klahr, 1999; Bullock & Sodian, 2003). Bei der Wahl der Aufgabenbeispiele fällt jedoch auf, dass die Experimentierbedingungen, unter denen die Forschungsergebnisse generiert werden, nicht verändert wurden. Ein Wechsel der Experimentierbedingungen kann z. B. realisiert werden, indem die Probanden die Gelegenheit erhalten, neben dem Gebrauch von vorgefertigten Versuchsmaterialien, ein eigenes Produkt zu erstellen (vgl. Kap. 2.3 ff.). In der Literatur sucht man allerdings vergeblich reale Experimentieraufgaben, die diese Anforderungen erfüllen, so dass gefragt werden muss, inwieweit die erhobenen Konzepte und Strategien für den Umgang mit Variablen als belastbare Ergebnisse eingestuft werden können, wenn ein Wechsel der Experimentierbedingungen vernachlässigt wurde. Diese Frage ist vor allem für den notwendigen Einsatz von Unterstützungsmaßnahmen beim Experimentieren interessant (vgl. Defizite), deren Wirksamkeit unter veränderten Experimentierbedingungen bisher leider nicht untersucht worden ist.

2.3 Entwicklung eines Aufgabenbeispiels zum offenen Experimentieren

Die Entwicklung des Aufgabenbeispiels zum offenen Experimentieren fußt auf den theoretischen Arbeiten zum offenen Experimentieren (Kap. 2.1) und zu den naturwissenschaftlichen Arbeitsweisen (Kap. 2.2).

2.3.1 Kriterien zur Entwicklung und Konstruktion des Aufgabenbeispiels

Auf die Entwicklung bzw. Konstruktion von Schülerexperimentieraufgaben haben ihre Zielsetzungen einen wesentlichen Einfluss. Vor allem die in der Literatur beschriebene Vielfalt der Ziele zeigt (Kap. 2.1.3), dass die Lernenden hohe Erwartungen und Anforderungen zu erfüllen haben, wenn sie eigenständig experimentieren. In der Literatur (Johnstone & Wham, 1982) wird die Vielfalt der Ziele kritisch bewertet, weil sie zu einer Überforderung der Experimentierenden führen können. Zu diesem Ergebnis kommt auch Hodson (1993), wenn er das Verhalten der Schüler mit „very busy getting nowhere“ beschreibt und folgert, dass

sich wesentliche Zielsetzungen von Schülerexperimenten wie learning science, learning about science und doing science nicht zusammen in einem experimentellen Ansatz verwirklichen lassen.

Bei der Entwicklung des eigenen Aufgabenbeispiels zum offenen Experimentieren werden die Ergebnisse der Autoren berücksichtigt, indem mit der offenen Experimentieraufgabe nur ein und nicht mehrere Hauptziele gleichzeitig verfolgt werden (Kriterium 1). Hierbei wird Abstand von dem Ansatz genommen, in einer Experimentieraufgabe gleichzeitig sowohl hohe fachinhaltliche Anforderungen, schwierige Methoden der Erkenntnisgewinnung als auch anspruchsvolle experimentelle Fertigkeiten vereinen zu wollen. Eine Anhäufung von verschiedenen Schwierigkeiten (inhaltliches Wissen, methodisches Wissen, strategisches Wissen), die zu einer Überforderung der Schüler führen können, soll mit einer Schwerpunktsetzung auf die naturwissenschaftlichen Arbeitsweisen am Beispiel des Umgangs mit Variablen vermieden werden.

Als weiteres Kriterium zur Entwicklung der offenen Experimentieraufgabe wird darauf geachtet, dass die Schüler eine relevante, schnell zu erfassende Fragestellung bearbeiten (Kriterium 2). Es soll verhindert werden, dass zur Vermittlung der Aufgabenstellung Problematisierungs- (s. u.) oder andere geleitete Hilfen notwendig werden. Als ein Beispiel für geleitete Hilfen beim eigenständigen Experimentieren seien Experimentieranleitungen genannt. Auf deren Einsatz wird verzichtet, weil sie unter anderem widersprüchliche Anforderungen zu erfüllen haben (Merzyn, 1994, S. 23): „Experimentieranleitungen sollen dem Schüler Anleitung und Hilfe bei eigenem Experimentieren geben. [...]. Sie sollen einerseits die Lesewilligkeit und Lesefähigkeit der Schüler nicht überfordern und sollen andererseits dennoch ins Einzelne gehen.“ Der Argumentation von Merzyn folgend, sind es aber nicht nur die widersprüchlichen Anforderungen, die an den Einsatz von Experimentieranleitungen gestellt werden, denn ein Abarbeiten vorgegebener Anweisungen kann sich ebenfalls als schwierig erweisen. Oftmals ist hierfür die Kenntnis von Fachtermini notwendig (Merzyn, 1987 & 1994, S.156), so dass schriftliche Experimentieranleitungen von den Lernenden nicht oder nur teilweise verstanden werden. Zum Verständnis der schriftlichen Anleitungen ist ein ausreichendes Leseverständnis erforderlich, damit beispielsweise Texte zum Erwerb experimenteller Routinen und Fertigkeiten erfolgreich eingesetzt werden können.

Obwohl der Einfluss des Leseverständnisses auf das (eigenständige) Experimentieren empirisch bisher wenig untersucht worden ist, sollen gerade Schüler mit einer geringen Lesekompetenz bei der Bearbeitung der offenen Experimentieraufgabe davon profitieren können, dass relevante Informationen der Aufgabenstellung und -bearbeitung nicht schriftlich dargeboten werden. Empirische Hinweise, dass das Leseverständnis einen Einfluss auf die Aufgabenbearbeitung haben kann, werden von Leutner am Beispiel des Problemlösens aufgezeigt (Leutner et al.,

2004). Die Gefahr eines Informationsverlustes, die bei einer nicht schriftlichen Präsentation der Aufgabenstellung gegeben ist, wird bei der Konstruktion der offenen Experimentieraufgabe durch eine kurze, schnell zu erfassende Fragestellung minimiert. Die Aufgabenformulierung kann sehr kurz gehalten werden, weil mit der Entwicklung des Aufgabenbeispiels zum offenen Experimentieren nur ein und nicht mehrere Hauptziele gleichzeitig verfolgt werden. Beispielsweise wird von einer Schwerpunktsetzung auf naturwissenschaftliche Arbeitsweisen nicht verlangt, dass mit dem Einsatz derselben Experimentieraufgabe auch komplizierte experimentelle Fertigkeiten vermittelt werden sollen, die häufig prozessbegleitende Anweisungen (verbal oder schriftlich) notwendig werden lassen.

Ein Verzicht auf Experimentieranleitungen ist jedoch noch kein Garant dafür, dass die Aufgabenstellung von den Lernenden schnell erfasst werden kann. Als Beispiel seien experimentell-offene Aufgabenstellungen von Reinhold (1996, 1997a) und Braun (2009) benannt, die mit der Auswahl und dem Einsatz komplexer (schwerer) physikalischer Phänomene ganz bewusst Konfliktsituationen provozieren möchten. Der Effekt, der auch als erklärungsbedürftige Situation oder produktive Störung bezeichnet wird, wird von den Autoren gezielt zum Einstieg in die Experimentieraufgabe benutzt, um bei den Adressaten sowohl Lehr- als auch Lernprozesse zu initiieren. „Der wesentliche Gesichtspunkt ist wohl, dass durch die Präsentation eines solchen Phänomens im Lehr-Lern-Prozess eine für die Beteiligten erklärungsbedürftige Situation entsteht.“ (Reinhold, 1996, S. 165); „Offenes Experimentieren beginnt in der Regel mit der Demonstration eines solchen Phänomens und versucht auf diese Weise eine erklärungsbedürftige Situation zu erzeugen. Dazu gehört ein gewisses Überraschungsmoment, ein Staunen.“ (Reinhold, 1997a, S. 46). Allerdings darf bei dieser Argumentationsfolge ihre Umkehrung nicht vernachlässigt werden. Demzufolge könnten beim Einstieg in das Experimentieren Lehr- und Lernprozesse verhindert werden, wenn sich die gewünschten Konfliktsituationen nicht einstellen. Von dieser Gefahr berichtet auch Reinhold und dokumentiert mit der Aussage „Für die beteiligten Studenten entstand keine erklärungsbedürftige Situation“ die nicht zu vernachlässigende Abhängigkeit des Effekts von der Auswahl der Phänomene (Reinhold, 1996, S. 165). Das Zitat bezieht sich auf ein komplexes (schwieriges) Phänomen, das aufgrund der Erfahrungen des Autors nicht zum Gegenstand seiner Fallstudien gemacht wurde. Die Defizite könnten minimiert, ggf. sogar kompensiert werden, indem den Experimentierenden zum Erkennen der erklärungsbedürftigen Situation Problematisierungshilfen angeboten werden. Es muss jedoch bedacht werden, dass deren Verwendung die Zugänglichkeit einer experimentellen Fragestellung erschweren würde. Aus diesem Grund werden bei der Konstruktion des eigenen Aufgabenbeispiels zum offenen Experimentieren keine Effekte in der Fragestellung verankert, die beim Einstieg in die Experimentiersituation eine produktive Störung intendieren. Dadurch soll vermieden werden, dass bereits zu

Beginn der Experimentierphase Lernschwierigkeiten aufgebaut werden, die eine vorzeitige Beendigung oder einen Abbruch der Experimentiertätigkeit bewirken können.

Angemerkt sei, dass zur Kennzeichnung des Aufgabenformats, das in den Studien von Reinhold (1996) und Braun (2009) benutzt wird, nicht der Begriff „offene Experimentieraufgabe“ sondern der Begriff „experimentell-offene Aufgabe“ verwendet wurde (s. o.). Die verbale Nuance soll betonen, dass die Autoren zur Lösung der Phänomene eine fachliche Erklärung anbieten, womit sie ermöglichen, die experimentell-offenen Experimentieraufgaben in experimentell-geschlossene Experimentieraufgaben zu überführen. Eine derartige Transformation kann unter dem Einfluss abänderbarer Experimentierphasen (Planung, Durchführung, Auswertung) vergleichsweise einfach realisiert werden. Dagegen wird die Transformation wesentlich schwerer, wenn eine experimentelle Fragestellung nicht nur von einer, sondern von mehreren Lösungen unterstützt wird. Der Einfluss abänderbarer Experimentierphasen wird dadurch erheblich reduziert. Um diesen Unterschied von Aufgabenformaten zu dokumentieren, wird zwischen den Begriffen „experimentell-offene Aufgabe“ und „offene Experimentieraufgabe“ unterschieden. Die zitierten Autoren stimmen darin überein, dass sie zu jedem Phänomen nur eine Erklärung anbieten. Zur Darstellung der Erklärungen werden physikalische Gleichungen, Modellierungen, Abbildungen oder Diagramme bemüht. Hinweise darauf, dass die Aufgabenstellungen mehrere Lösungen unterstützen, werden aus den Arbeiten von Reinhold (1996) und Braun (2009) nicht ersichtlich.

Mit dem dritten und letzten Kriterium zur Entwicklung der offenen Experimentieraufgabe werden die geöffneten Kategorien des Aufgabenbeispiels bestimmt (Kriterium 3). Die geöffneten Komponenten werden fixiert, indem folgende Fragen beantwortet werden: Was wird geöffnet?, Wie wird geöffnet?, Wann wird geöffnet?. Da die differenzierten Fragen nicht nur der zielgerichteten Konstruktion, sondern auch der zielgerichteten Verortung der offenen Experimentieraufgabe im eigenen Ansatz des offenen Experimentierens dienen (vgl. Kap. 2.3.3), wird das dritte Kriterium zur Entwicklung des Aufgabenbeispiels im Anschluss der Präsentation und Diskussion des Aufgabenbeispiels im Kapitel 2.3.3 vorgestellt.

2.3.2 Präsentation und Diskussion des entwickelten Aufgabenbeispiels

Den Adressaten wird folgende offene Experimentieraufgabe angeboten:

Konstruiere mit dem bereitgestellten Experimentiermaterial die Rotorblätter für das Modell eines Windrades so, dass bei den vorgegebenen Versuchsbedingungen (Fön und dessen Abstand zum Windrad) eine möglichst hohe Spannung erzeugt wird.

Zur Lösung des Aufgabenbeispiels ist den Lernenden die Gestaltung der Rotoren vollständig selbst überlassen. Alle notwendigen Materialien werden zur Verfügung gestellt. Diese umfassen u. a. den auf einen Standfuß montierten Generator, ein Voltmeter, einen Fön sowie Geräte zur Fertigung und Erwärmung des thermoplastischen Flügelmaterials. Die gebauten Windrotoren müssen an dem Generator verschraubt (vgl. Anhang) und mit dem Fön, der ohne Heizfunktion betrieben wird, umströmt werden. Die erzeugte Spannung wird mit dem Voltmeter im stromlosen Zustand gemessen (Priemer & Kirchner, 2007a).

Um zu sinnvollen Lösungen zu gelangen, werden den Experimentierenden Hilfsmittel, wie der Zugriff auf das Internet und schriftliche Informationen aus Lehrbüchern, zur Verfügung gestellt. Weiterhin besteht die Möglichkeit, vorgefertigte Sets von Windradprofilen (Synonym: Flügelsets) zu nutzen. Das aufgabenspezifische Experimentiermaterial zeichnet sich dadurch aus, dass jeweils nur ein Parameter der Flügelform verändert wurde. Die systematischen Messfolgen bieten sich an, um den Einfluss der Parameter auf die Spannung des Windrades zu untersuchen. Insgesamt werden acht verschiedene Flügelsets bereitgestellt, mit denen der Einfluss der Länge, der Breite, des Anstellwinkels, der Wölbung, von Winglets, der (Grund-)Form, der Verdrehung (Verschränkung) und der Flügelanzahl getestet werden kann. Da die Messungen als Grundlage für die eigenen Flügelkonstruktionen dienen, wurden in den Sets nicht nur Variablen mit linearen und extremalen Verlauf, sondern auch Faktoren, die die Spannung nicht beeinflussen mit aufgenommen. Weiterhin bestand die Absicht, mit den Flügelsets nicht alle aus der Theorie bekannten, einflussnehmenden Parameter zu bedienen. Die Lernenden sollen so die Möglichkeit erhalten, mit eigenen Experimenten nach weiteren spannungsabhängigen Variablen zu suchen. Grundsätzlich dürfen hierfür auch Trial- & Error-Verfahren angewendet werden, die zu kreativen Ideen und systematischen Untersuchungen führen können. Als Beispiel einer Variablen, von der bekannt ist, dass sie den Auftrieb an Tragflächenprofilen verändert und die nicht als Messfolge der Flügelsets zur Verfügung gestellt wird, sei die Oberflächenbeschaffenheit des Profilmaterials genannt (Dubs, 1966, S. 108 ff.). Mit kreativen Ideen und eigenen Experimenten könnte in einer Untersuchung mit Kontrolldesign die Rauigkeit der Flügeloberseite, die Rauigkeit der Flügelunterseite oder der beidseitig aufgeraute Flügel getestet werden. Weiterhin könnte der Verlauf der Variablen bestimmt werden, wenn auf den korrespondierenden Flügelseiten die Ausprägung der Rauigkeit systematisch verändert wird. Abschließend sei zu den Messfolgen der Flügelsets angemerkt, dass jedes Set mehrmals reproduziert wurde. Zudem enthält jedes Set einen Referenzflügel, der sich durch identische Parameter auszeichnet. Beide Maßnahmen eröffnen die Möglichkeit, dass die Flügelsets nicht nur in der dargebotenen Zusammenstellung benutzt werden können, sondern auch zur Kombination weiterer Messfolgen einladen, mit denen die Abhängigkeit von Profil-

variablen unter dem Einfluss veränderter Ausgangsparameter getestet werden kann. Beispielsweise gestattet die Konzeption, dass die Variable Flügellanzahl mit verschiedenen Anstellwinkeln untersucht werden kann.

Das entwickelte Aufgabenbeispiel zeigt, dass die Adressaten bei der Bearbeitung der offenen Experimentieraufgabe zahlreiche Gelegenheiten für den Umgang mit Variablen haben. Die intendierten Experimentiertätigkeiten und -prozesse der Lernenden könnten schon vor Beginn der empirischen Erhebung mit einem Erwartungshorizont schriftlich fixiert werden. Da mit dieser Studie ein explorativer Ansatz verfolgt wird, werden die Tätigkeiten der Experimentierenden zuvor nicht durch normative Erwartungen detailliert festgelegt.

In dem Kapitel 2.3.1 wird herausgestellt, dass der Schwerpunkt des entwickelten Aufgabenbeispiels bei den naturwissenschaftlichen Arbeitsweisen für den Umgang mit Variablen liegt. Es wird nicht verlangt, dass mit dem Einsatz der offenen Experimentieraufgabe auch der anspruchsvolle Fachinhalt, zu dem u. a. das Prinzip des dynamischen Auftriebs zählt, vermittelt wird. Trotzdem darf nicht vergessen werden, dass strategische Entscheidungen der Lernenden von ihren fachlichen Kenntnissen moderiert werden können. Um diesen Effekt zu nutzen, wird die offene Experimentieraufgabe in ein Projekt eingebunden (vgl. Kap. 3). In der Projekteinführungsphase werden fachliche Inhalte, wie das Prinzip des dynamischen Auftriebs, verankert. Dadurch wird gewährleistet, dass die hohen fachlichen Anforderungen des Themas von der offenen Experimentieraufgabe entkoppelt und dennoch thematisiert werden. Weiterhin erfüllt die Vermittlung fachlicher Inhalte in der Projekteinführungsphase die Funktion, das Vorwissen der Zielgruppe zu synchronisieren. Eine Besonderheit des entwickelten Aufgabenbeispiels ist, dass mit der Fragestellung nur ein und nicht mehrere Hauptziele gleichzeitig verfolgt werden. Exemplarisch soll zum Abschluss des Kapitels aufgezeigt werden, wie durch eine Variation der (Mess-)Methoden, die intendierten Hauptziele der Experimentieraufgabe verändert werden können. Beispielsweise könnte zur Dimensionierung der Profilkonstruktionen anstelle des Spannungsmessgerätes ein Stroboskop eingesetzt werden. Diese Maßnahme hätte zur Konsequenz, dass zur Bearbeitung der offenen Experimentieraufgabe neben den strategischen Wissensaspekten mehr (mess-) methodische Wissensaspekte notwendig würden, weil die Synchronisation der Blitzfrequenz mit den rotierenden Windradprofilen als ein komplexeres (schwereres) Messverfahren beurteilt wird als die Spannungsmessung mit einem konventionellen Voltmeter. Die Schwierigkeiten, die mit dem Einsatz eines Stroboskops verbunden sind, liegen nicht nur in der möglichst genauen Justierung der Blitzfrequenz (experimentelle Fertigkeiten zur Verringerung von Messunsicherheiten) oder in der Vermittlung des anspruchsvolleren Messprinzips. Auch die Frage, welche Blitzfrequenz bzw. welcher Messwert der Umlaufgeschwindigkeit des luftumströmten Windrades entspricht, ist für die Lernenden nicht einfach zu

beantworten, denn: Mit der sukzessiven Erhöhung der Blitzfolge entstehen im Kegel der Lichtquelle mehrere, scheinbar synchronisierte „Standbilder“ der Windradprofile, die einen Abgleich des Stroboskoplichtes mit der Umlaufgeschwindigkeit der Windradflügel vortäuschen.

Im Gegensatz zu der entwickelten Variante des Aufgabenbeispiels mit Spannungsmessung, kann der Generator des Windrades auch mit einem elektrischen Verbraucher (Ohmscher Widerstand, Motor) belastet werden. Die Maßnahme hätte zur Konsequenz, dass die Profile der Windräder solange verändert werden müssten, bis die maximale elektrische Ausgangsleistung vom Generator des Windradmodells abgenommen werden könnte. Zur Bearbeitung dieser Variante der offenen Experimentieraufgabe sind – gemessen an der Datenaufnahme mit einem Spannungsmesser – neben den strategischen Wissensaspekten auch umfangreichere inhaltliche und (mess-)methodische Wissensaspekte notwendig. Zu den inhaltlichen Wissensaspekten zählt nicht nur die Kenntnis der physikalischen Gleichung der elektrischen Leistung. Ebenso muss bedacht werden, dass die elektrische Leistung mit der Wahl des Verbrauchers (Ohmscher Widerstand) variiert. Infolgedessen würden die Profilkonstruktionen in Abhängigkeit des elektrischen Verbrauchers dimensioniert. Das bedeutet, dass durch einen Austausch der Ohmschen Widerstände ggf. die eingestellten Parameter der Rotorflügel modifiziert werden müssen, wenn weiterhin vom Generator des Windradmodells die maximale Leistung abgenommen werden soll. Um den Experimentierenden diesen Sachverhalt zu kommunizieren, bietet sich an, ein selbst gefertigtes Flügelprofil systematisch durch Variation von Ohmschen Widerständen zu vermessen (Priemer & Kirchner, 2007a). Damit diese Aspekte vermittelt werden können, sind zur Bearbeitung der Experimentieraufgabe neben den strategischen und inhaltlichen Wissensaspekten auch (mess-)methodische Wissensaspekte notwendig, denn: Der fachgemäße Aufbau und Betrieb eines elektrischen Schaltkreises, der einen Generator, einen Ohmschen Widerstand, einen Strommesser und einen Spannungsmesser enthält, ist nach den Erfahrungen der Unterrichtspraxis für Lernende auch nach Abschluss der Sekundarstufe I mit Schwierigkeiten verbunden, die ohne Einsatz von Unterstützungsmaßnahmen (Experimentieranleitungen, Schaltskizzen o. ä.) nicht bewältigt werden können. Mit dieser Variante des Aufgabenbeispiels soll aufgezeigt werden, wie durch die Aufnahme eines elektrischen Verbrauchers in den Schaltkreis des Generators mehrere Schwierigkeiten (inhaltliches Wissen, strategisches Wissen, methodisches Wissen) in einer Experimentieraufgabe vereint werden, so dass auf eine Schwerpunktsetzung der Hauptziele verzichtet werden muss.

2.3.3 Verortung des Aufgabenbeispiels im eigenen Ansatz des offenen Experimentierens

Damit das entwickelte Aufgabenbeispiel im eigenen Ansatz des offenen Experimentierens verortet werden kann, bedarf es einer Positionierung hinsichtlich der Fragen Welche Komponenten der offenen Experimentieraufgabe werden geöffnet? (Dimensionen der Öffnung) und Wie stark werden die Komponenten der offenen Experimentieraufgabe geöffnet? (Grade der Öffnung). Beide Fragestellungen werden beantwortet, indem ein Bezug des Aufgabenbeispiels zu den Dimensionen und Öffnungsgraden hergestellt wird. Nach dem eigenen Ansatz wird die Offenheit von Experimentieraufgaben durch drei Dimensionen abgebildet. Damit das entwickelte Aufgabenbeispiel in dem Ansatz verortet werden kann, müssen alle Dimension mit den zugehörigen Teildimensionen erklärt werden. Die Erklärungen werden von den leitenden Fragen Was wird geöffnet? (Leitfrage der 1. Dimension), Wie wird geöffnet? (Leitfrage der 2. Dimension) und Wann wird geöffnet? (Leitfrage der 3. Dimension) begleitet. Zur ersten Dimension, die den Fachinhalt, die Strategie und die Methode offener Experimentieraufgaben abbildet, gehören die Teildimensionen Fachinhalt (1a), Strategie (1b) und Methode (1c). Zur zweiten Dimension, die die Lösungen und Lösungswege offener Experimentieraufgaben abbildet, gehören die Teildimensionen Lösungen (2a) und Lösungswege (2b). Zur dritten Dimension, die die Phasen des Experimentierens abbildet, gehören die Teildimensionen Aufgabenstellung (3a), Planungsphase (3b), Durchführungsphase (3c) und Auswertungsphase (3d).

1. Dimension: Was wird geöffnet?

1a) Fachinhalt. Mit der konzipierten Experimentieraufgabe werden die fachlichen Inhalte Windenergie, dynamischer Auftrieb, Generatorprinzip und elektrische Spannung fokussiert. Eine Offenheit bezüglich der Wahl des Fachinhaltes bietet das Aufgabenbeispiel nicht. Die Graduierung ist in dieser Klassifizierung bereits enthalten, da zwischen den drei Stufen „der Fachinhalt wird genau vorgegeben“, „mehrere Fachinhalte werden zur Wahl gestellt“ und „der Fachinhalt kann ohne Vorgaben gewählt werden“ differenziert wird. (Graduierung der Teildimension Fachinhalt: vorgegeben).

1b) Strategie. Das Aufgabenbeispiel unterstützt eine strategische Offenheit, denn es werden keine Vorgaben zu der Herangehensweise gemacht, wie die Experimentieraufgabe gelöst werden soll. Durch die strategische Offenheit des Aufgabenbeispiels werden den Experimentierenden beispielsweise die Festlegung des Versuchsdesigns, die Verfahren und Strategien zur Aufnahme und Auswertung der Messdaten, die Festlegung notwendiger Experimentierschritte und die Anzahl der aufzunehmenden Messwerte zur Wahl gestellt.

Es sei angemerkt, dass mit der Bereitstellung des vorgefertigten, aufgabenspezifischen Experimentiermaterials Flügelsets die strategische Offenheit auch vorskizziert werden kann, indem vor Aufnahme der Experimentiertätigkeiten z. B. die Strategie der Variablenkontrolle explizit unterrichtet wird. Jedoch muss bedacht werden, dass aus den bereitgestellten Flügelsets neue Messfolgen zusammengestellt werden können, so dass den Adressaten die Möglichkeit eröffnet wird, die Abhängigkeit einer Variablen mit anderen Ausgangsparametern zu untersuchen. Beispielsweise kann die Variable Flügelanzahl mit unterschiedlichen Anstellwinkeln oder mit unterschiedlichen Flügelbreiten getestet werden. Die Option birgt die Gefahr, dass aus den Profilen der Flügelsets neue Messfolgen kombiniert werden, die einem konfundierenden Versuchsdesign folgen. (Graduierung der Teildimension Strategie: keine Vorgaben).

1c) Methode. Um die Umlaufgeschwindigkeiten der Windräder zu bestimmen, werden kommerzielle Spannungsmesser benutzt. Da die Entscheidung über den Einsatz der Messmethode von den Experimentierenden nicht selbst getroffen werden kann, wird eine methodische Offenheit von diesem Aufgabenbeispiel nicht unterstützt. (Graduierung der Teildimension Methode: vorgegeben).

2. Dimension: Wie wird geöffnet?

Die Komplexität – im Sinne von Vielfalt einer Experimentieraufgabe – wird von der Anzahl der möglichen Lösungen (eine oder mehrere) und der Anzahl der möglichen Lösungswege (einer oder mehrere) bestimmt. Bei dem entwickelten Aufgabenbeispiel handelt es sich um eine komplexe offene Experimentieraufgabe, die mehrere Lösungen unterstützt, die auf verschiedenen Lösungswegen erreicht werden können.

2a) Anzahl der Lösungen. Die Konstruktion von Rotorblättern für das Modell einer Windenergieanlage mit möglichst hoher Spannung lässt unterschiedliche Lösungen zu. Es können mehrere Profilkonstruktionen für die Windradmodelle gefunden werden, die sich in ihrer geometrischen Form grundlegend unterscheiden und unter denselben Anströmverhältnissen vergleichbar hohe Spannungen (Drehzahlen) erzielen. Die Experimentierenden können aber nicht nur an den selbst erstellten, sondern auch an den bereitgestellten Profilformen der Flügelsets erkennen, dass das Aufgabenbeispiel mehrere Lösungen haben muss. Obwohl die Vielzahl der Lösungen gegenüber den eigenen Profilkonstruktionen deutlich reduziert worden ist, befinden sich unter den Messfolgen der Flügelsets drei Profile verschiedener Bauart, die annähernd gleich hohe Spannungen (Drehzahlen) erreichen. Zudem können zwei Profile identifiziert werden, die sich nur in der Ausprägung eines Parameters unterscheiden, aber unter allen Profilen der Flügelsets den Spannungsspitzenwert (Drehzahlspitzenwert) einnehmen.

Die Existenz mehrerer Lösungen kann durch systematische Arbeiten an umströmten Tragflächenmodellen verdichtet werden, die nach ihrer funktionalen Beschreibung konstruiert wurden (Glomski, 2008). Mit der zitierten Untersuchung wurde die Frage verfolgt, welchen Einfluss eine systematische Variation der Profilparameter Nasenradius, Dicke, Dickenrücklage, Wölbung und Wölbungsrücklage auf die Auftriebs- und Widerstandsbeiwerte verschiedener Profilgrundformen hat. Ein Vergleich von 45 Polardiagrammen führte zu dem Ergebnis, dass die maximalen Auftriebsbeiwerte von mehreren Tragflächenprofilen erreicht werden. Die Geometrien der korrespondierenden Flügel waren grundlegend verschieden. Ein weiteres Ergebnis der Untersuchung war, dass in Analogie zu den Profilen mit maximalen Auftriebsbeiwerten mehrere Flügelformen bestimmt werden konnten, die minimale Widerstandsbeiwerte haben. Die Ergebnisse machen deutlich, dass die Auswahl günstiger Profilformen sehr stark von der Untersuchungsfrage dominiert wird. Zu diesem Ergebnis führen auch die Versuche mit den Windradmodellen, an denen aufgezeigt werden kann, dass die Parameter der Rotorblätter unter dem Einfluss wechselnder elektrischer Verbraucher modifiziert und neu dimensioniert werden müssen (ohne Lastwiderstand: Messung der maximalen Spannung). Ein Vergleich mit den Untersuchungen an den Tragflächenmodellen legt nahe, dass das entwickelte Aufgabenbeispiel selbst bei einem potentiellen Wechsel der Aufgabenstellung (Maximierung der Spannung, Maximierung der elektrischen Leistung) immer noch von mehreren Lösungen getragen wird. (Graduierung der Teildimension Lösung: mehrere Lösungen).

2b) Anzahl der Lösungswege. Die Lösungen des Aufgabenbeispiels können auf mehreren Lösungswegen erreicht werden. Beispielsweise kann die Lösungsfindung durch theoretische Hintergründe und praktische Erfahrungen, durch systematische eigene Experimente, durch Recherchen im Internet und Literatur und durch Trial-& Error-Verfahren initiiert und vollzogen werden. (Graduierung der Teildimension Lösungswege: mehrere Lösungswege).

3. Dimension: Wann wird geöffnet?

3a) Fragestellung. Die Aufgabenstellung der Experimentieraufgabe wird vorgegeben. Den Experimentierenden wird beispielsweise nicht zur Wahl gestellt, ob sie einen Auftriebsläufer oder einen Widerstandsläufer konstruieren und, ob die Profile des Auftriebsläufers eine möglichst hohe Spannung (Drehzahl) oder eine möglichst hohe Leistung erreichen sollen. (Graduierung der Teildimension Aufgabenstellung: vorgegeben).

3b) Planung. 3c) Durchführung. 3d) Auswertung. Die drei Phasen des Experimentierens sind offen, denn den Experimentierenden werden keine Vorgaben zum Vorgehen gemacht. Es besteht die Freiheit, über die Sequenzierung und über die Länge der Experimentierphasen zu entscheiden, in denen die Kategorien Fachinhalt-

Strategie-Methode (1. Dimension) bzw. Lösungen und Lösungswege (2. Dimension) verankert werden. Die Planungen der Lernenden können z. B. Strategien zur Lösungsfindung und zur Durchführung der Experimente umfassen. Darüber hinaus werden Hilfsmittel wie der Zugriff auf das Internet, schriftliche Informationen aus Lehrbüchern und vorbereitetes Experimentiermaterial (Flügelsets) zur Verfügung gestellt, auf das wahlweise zugegriffen werden kann. Da auch für die Phasen der Durchführung und der Auswertung keine Vorgaben gemacht werden, liegt es im Rahmen der Möglichkeiten, dass die selbst hergestellten Profile nach der eigenen Intuition angefertigt werden oder aus der systematischen Arbeit mit den Flügelsets hervorgehen. Denn nicht nur die Fertigung der eigenen Profile, sondern auch die Arbeit mit den Flügelsets bietet den Experimentierenden die Möglichkeit, die Phasen der Planung, der Durchführung (z. B. Art und Umfang der Datenaufnahme, Protokoll der Messergebnisse) und der Auswertung (z. B. grafische Visualisierung der Messergebnisse, Vergleich und Bewertung von Messfolgen) selbständig zu gestalten. (Graduierung der Teildimensionen Planung, Durchführung und Auswertung: keine Vorgaben zum Vorgehen).

2.3.4 Kritische Anmerkungen zur Entwicklung des Aufgabenbeispiels

Offene Experimentieraufgaben können u. a. durch die Vielfalt ihrer Lösungen (eine oder mehrere) und Lösungswege (eine oder mehrere) unterschieden werden (Kap. 2.3.3). Der Stellenwert dieser Aufgaben wurde durch bildungspolitische Programme wie durch das Programm der Bund-Länder-Kommission (BLK) zur Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts (SINUS) stark verändert. Trotzdem darf für die naturwissenschaftlichen Fächer konstatiert werden, dass den Lernenden noch zu wenige offene Experimentieraufgaben angeboten werden, die von mehreren Lösungen getragen werden. Eine mögliche Ursache könnte die Schwierigkeit ihrer Konzeption sein, was an ausgewählten Aufgabenbeispielen aus dem BLK-Programm SINUS aufgezeigt werden soll. Hier führte die Öffnung von (Experimentier-)Aufgaben zu Lernsettings, die von mehreren Lösungswegen getragen werden (Fröhlich, Bieber & Horn, 2004). Dazu haben die Autoren Aufgabenbeispiele benutzt, deren Lösung auf einer physikalischen Gleichung beruht. An den geöffneten Varianten der Aufgabenbeispiele kann jedoch nicht erkannt werden, dass die Transformation zu einer erhöhten Anzahl der möglichen Lösungen geführt hat. Mit Bezug auf die Aufgabenbeispiele des BLK-Programms wird vermutet, dass offene Experimentieraufgaben mit mehreren Lösungen nicht durch den Öffnungsprozess experimentell-geschlossener Aufgaben entstehen. Stattdessen wird propagiert, dass offene Experimentieraufgaben mit

mehreren Lösungen aus kriteriumsgeleiteten respektive dimensionsgeleiteten Entwicklungsprozessen erwachsen.

Trotz des Mangels an Angeboten von offenen Experimentieraufgaben mit mehreren Lösungen, muss deren Einsatz kritisch diskutiert werden. Kritiker mögen behaupten, dass Lernende überfordert werden, wenn Experimentieraufgaben zahlreiche Lösungen zulassen. Hierbei darf die Umkehrung der Argumentation nicht vernachlässigt werden, wonach das entwickelte Aufgabenbeispiel der offenen Experimentieraufgabe dazu beiträgt, dass Lernende von der Vielfalt der Lösungen profitieren können. Denn anders als bei Aufgabenbeispielen mit nur einer Lösung werden die Adressaten bei der Lösungsfindung nicht genötigt, die eine existierende Lösung (wieder-) zuentdecken. Zugleich muss bedacht werden, dass mit dem Ansatz der Schwerpunktsetzung nicht verschiedene Schwierigkeiten in einer Experimentieraufgabe angehäuft werden (inhaltliches Wissen, strategisches Wissen, methodisches Wissen). Der Schwerpunkt des Aufgabenbeispiels liegt bei den naturwissenschaftlichen Arbeitsweisen. Um eine Überforderung der Lernenden zu verhindern, sind (fach-)inhaltliche und (mess-)methodische Wissensaspekte reduziert worden. Die konzipierte offene Experimentieraufgabe dient als Aufgabenbeispiel in den explorativen Vorstudien (Kap. 3) und in der Hauptstudie (Kap. 4). Sowohl in den Vorstudien als auch in der Hauptstudie wird die offene Experimentieraufgabe in ein Projekt eingebunden.

Empirischer Teil

3 Explorative Vorstudien

In diesem Kapitel werden zwei aufeinander aufbauende explorative Vorstudien präsentiert. Das Interesse der ersten explorativen Vorstudie liegt in der Analyse von Planungsprozessen, die Probanden für den Umgang mit Variablen bei der Bearbeitung einer offenen Experimentieraufgabe zeigen. Durch die Analyse sollen Defizite in den Planungsprozessen herausgestellt werden (Kap. 3.1). Zur Reduktion bzw. Kompensation der Probleme und Defizite wird in der zweiten explorativen Vorstudie und der Hauptstudie folgender didaktischer Ansatz gewählt: Vor der Bearbeitung der offenen Experimentieraufgabe erhalten die Probanden eine einmalige, einführende, informierende, explizite und vorbereitende Instruktion. Die Umsetzung der vorbereitenden Instruktion erfolgt durch Unterricht, der auf einem lernzielorientierten Konzept für den Umgang mit Variablen aufbaut (Kap. 3.2). Mit der zweiten Vorstudie soll die Wirksamkeit des didaktischen Ansatzes hinsichtlich des planerischen und experimentellen Umgangs mit Variablen, den die Probanden bei der Bearbeitung der offenen Experimentieraufgabe zeigen, exploriert werden (Kap. 3.3). Die Wirksamkeitsprüfung des didaktischen Ansatzes ist ein Anliegen der Hauptstudie, was u. a. die Entwicklung eines adäquaten Messinstruments notwendig macht. Das in der Hauptstudie benutzte Messinstrument Beobachtungsbogen wird aufgrund differenziert beobachteter Experimentiersequenzen als Ergebnis der ersten und zweiten Vorstudie erstellt (vgl. Kap. 3.3 und Kap. 4.2.3). Nach einer kritischen Reflexion der Vorstudien (Kap. 3.4) soll mit einem weiteren Ausblick auf die Hauptstudie berichtet werden, dass die Wirksamkeit verschiedener Ausprägungen des Vermittlungskonzeptes zum Umgang mit Variablen geprüft wird. Die Ausprägungen des Vermittlungskonzeptes, die zum Abschluss des dritten Kapitels (Kap. 3.5) vorgestellt und verglichen werden, unterscheiden sich durch ihre (Erwerbs-) Kontexte, wohingegen die intendierten Lernziele und der didaktisch-methodische Aufbau des Vermittlungskonzeptes für alle Ausprägungen identisch sind.

3.1 Explorative Vorstudie I

Das Kapitel 3.1 wird durch die drei Abschnitte Forschungsfragen, Methoden und Ergebnisse strukturiert. Folgende *Forschungsfragen* werden mit der explorativen Vorstudie I untersucht:

1. Welche Planungsschritte zeigen Schüler hinsichtlich des Umgangs mit Variablen bei einer offenen Experimentieraufgabe?
2. Welche Schwierigkeiten der Schüler sind für den Umgang mit Variablen bei den Planungsschritten einer offenen Experimentieraufgabe festzustellen?

Methoden. Design. Die Fragestellungen werden am Beispiel der Profilkonstruktion für das Modell eines Windrades untersucht (Kap. 2.3, s. u.). Dem Design der Abbildung 2 ist zu entnehmen, dass den Probanden im Rahmen der Vorstudie I vor und während der Bearbeitung der offenen Experimentieraufgabe keine Instruktionen oder unterstützenden Hilfen, wie z. B. rezeptive Versuchsanleitungen, gestufte Hilfen oder Prompts, angeboten werden.

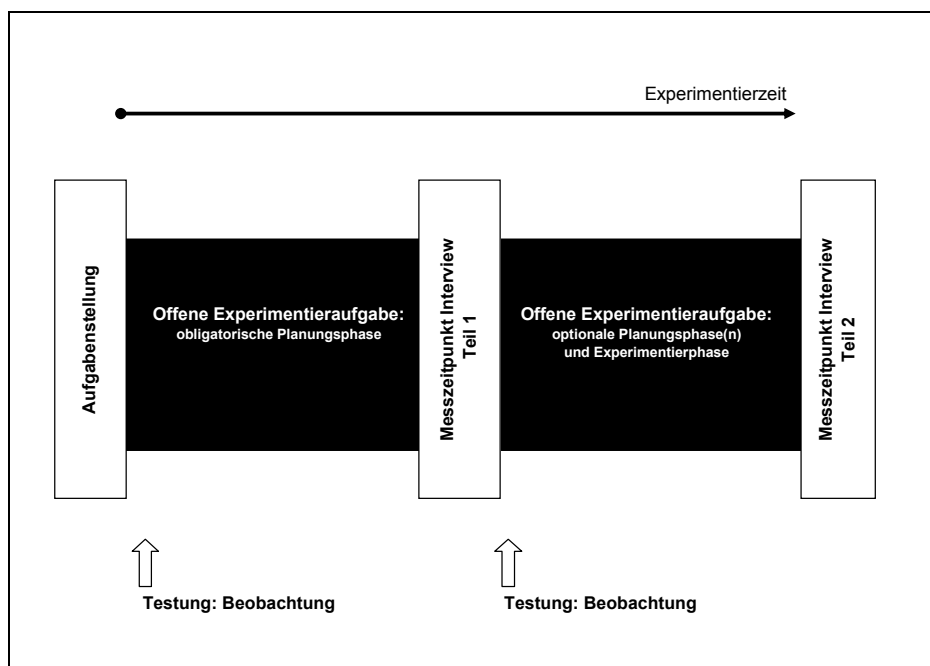


Abb. 2: Design der explorativen Vorstudie I

Die Datenerhebung erfolgt durch gezielte Beobachtungen der Planungsschritte, durch die Analyse des selbst erstellten schriftlichen Materials der Probanden und durch halbstrukturierte Interviews. Um nicht nur die ersten Planungsschritte im Anschluss der Aufgabenstellung, sondern auch die Änderungen der Planungsschritte durch die Phase des Experimentierens zu erfassen, werden mit dem Messinstrument

Interview die Erhebungsdaten an zwei Messzeitpunkten generiert. Das erste Interview findet nach einer obligatorischen, maximal 20 Minuten dauernden Versuchsplanungsphase statt. Das zweite Interview wird nach Vollendung der maximal 150 Minuten dauernden Experimentierphase geführt. Dem Design ist nicht zu entnehmen, dass die offene Experimentieraufgabe von allen Probanden in den Fachräumen des Schülerlabors der Ruhr-Universität Bochum (Alfried-Krupp Schülerlabor, AKS) bearbeitet wird, welcher auch für die Hauptstudie als Ort der Datenerhebung mit Kontrolle der Experimentierbedingungen und mit Kontrolle der Experimentierumgebung (Schülerlaborvariablen) dient (Kap. 4.2.1). Im AKS werden die Planungsschritte und -defizite für den Umgang mit Variablen beim Experimentieren erhoben. Alle Kerndaten des Untersuchungsdesigns werden mit der Abbildung 2 zusammengefasst.

Stichprobe. Das Forschungsinteresse der ersten Vorstudie liegt bei Probanden, die Physikunterricht in den Jahrgangsstufen 8 bis 10 der Sekundarstufe I an Gymnasien in Nordrhein-Westfalen erhalten haben (G 9), ein hohes fachliches und thematisches Interesse mitbringen und gute Noten in Physik haben. Aufgrund der genannten Merkmale darf der Stichprobe unterstellt werden, dass die explorierten und deskriptiv generierten Ergebnisse bzgl. des Umgangs mit Variablen beim Planen von Experimenten den oberen Niveaubereich der Zielgruppe markieren. Um gleichzeitig mögliche jahrgangsstufenspezifische Unterschiede zu erfassen, wird eine altersheterogene Stichprobe gewählt, bei deren Ziehung eine Gleichverteilung der Probanden über die Jahrgangsstufen beachtet wird. Damit die Auswahlkriterien garantiert werden können, werden Probanden rekrutiert, die sich aus dem regionalen und überregionalen Umfeld des Schülerlaborstandortes auf ein Ferienangebot der Fakultät für Physik und Astronomie der Ruhr-Universität Bochum beworben haben. Die Auswahl der Bewerber erfolgt zufällig. Da es sich nicht um eine Klumpenstichprobe (Schulklassen) handelt, können klassen- und schulspezifische Effekte ausgeschlossen werden. An der ersten explorativen Vorstudie nehmen insgesamt 18 Probanden teil, die für die Bearbeitung der im Kapitel 2.3 vorgestellten offenen Experimentieraufgabe zu neun altershomogenen Probandenpaaren (Partnerarbeit) zusammengeführt werden.

Offene Experimentieraufgabe. Für die Probanden wird die zu lösende Experimentieraufgabe „Konstruiere die Rotorblätter für das Modell eines Windrades so, dass bei den vorgegeben Versuchsbedingungen (Fön und dessen Abstand zum Windrad) eine möglichst hohe Spannung erzeugt wird“ methodisch in ein Projekt eingebunden. Der Projektablauf wird durch folgende drei Phasen gegliedert: Einführungsphase, Experimentierphase und Präsentationsphase. Eine ausführliche Projektbeschreibung befindet sich in der Literatur (Priemer & Kirchner, 2007a).

In der Einführungsphase des Projektes erhalten die Probanden durch einen Lehrer-experimentalvortrag theoretische Basisinformationen. Die Inhalte wurden unter Berücksichtigung der Literatur ausgewählt (Korneck, 1998; Wodzinski, 1999; Wodzinski & Ziegler, 2000; Weltner, 2000, 2002). Dazu gehören zum einen physikalische Grundlagen wie Energieumwandlungsprozesse an Windkraftanlagen, Typisierungen von Windrädern (Auftriebs- und Widerstandsläufer, mit Experimenten), der Bernoulli-Effekt (mit Experimenten), das Prinzip des dynamischen Auftriebs (mit Experimenten) und das Generatorprinzip (mit Experimenten). Zum anderen werden zu den Basisinformationen auch meteorologische Aspekte (Windgeschwindigkeiten und -stärke) und energiepolitische Ergänzungen (z. B. Energiebereitstellung) gezählt. Die Phase wird mit der Aufgabenstellung und der Einführung in die Experimentiermaterialien beendet. In der Experimentierphase, zu der nicht nur die Konstruktion der Lösungen, sondern auch die Planung der Eigenkonstruktionen gezählt wird, wird von den Probandenpaaren die offene Experimentieraufgabe bearbeitet. Das vorgefertigte, aufgabenspezifische Experimentiermaterial Flügelsets sowie der Zugriff auf das Internet werden in dieser Phase als Hilfsmittel angeboten. Der Projektabschluss wird mit der Präsentationsphase eingeleitet. Im Zentrum der Phase liegt der qualitative und quantitative Vergleich der selbst gebauten Windradmodelle, wozu die Probandenpaare in eine Wettbewerbssituation versetzt werden. Mit einem Messstand werden die maximalen Spannungen der Eigenkonstruktionen unter den gleichen, vorher festgelegten Bedingungen geprüft.

Zur Einführungsphase des Projektes muss angemerkt werden, dass naturwissenschaftliche Arbeitsweisen und fachspezifische Methoden, wie der Umgang mit Variablen beim Experimentieren, nicht zu den transportierten Basisinformationen des Lehrerexperimentalvortrags zählen. Insbesondere die Vermittlung des Variablenkonzeptes ist ausschließlich der auf das Projekt bzw. der auf die offene Experimentieraufgabe vorbereitenden Instruktion vorbehalten (Kap. 3.2). Andernfalls wäre die Unterscheidbarkeit zwischen den Experimentalgruppen nicht gewährleistet.

Messinstrumente. Die Planungsschritte der Probanden (Untersuchungsgegenstand) werden mit mehreren Messmethoden erhoben. Zur Auswertung liegen die gezielten Beobachtungen aus qualitativen Beobachtungsprotokollen, die mündlichen Befragungen aus halbstrukturierten Leitfadeninterviews sowie das schriftliche Material aus den selbst erstellten Dokumentationen der Probanden (schriftliche Selbstauskunft) vor. Die Anwendung mehrerer Messmethoden (Methoden-triangulation) wird gewählt, damit der Untersuchungsgegenstand empirisch-qualitativ umfassend exploriert werden kann. Die angewandten Verfahren unterscheiden sich durch den Grad ihrer Reaktivität, wobei das selbst erstellte schriftliche Material der Probanden als nicht reaktive Datenerhebungsmethode

eingestuft wird. In Übereinstimmung mit den Forschungszielen wird eine freie, offene, unstandardisierte, qualitative Beobachtungstechnik verwendet, so dass die Beobachtungsprotokolle eine Dokumentation der beobachtbaren Handlungen, Experimentiertätigkeiten und den verbalen Argumentationen (verbale Selbstauskünfte der Probandenpaare) enthalten. Es sei insbesondere darauf hingewiesen, dass die Beobachtungsprotokolle noch nicht mit Hilfe eines strukturierten Beobachtungsbogens erstellt wurden, denn: dessen Entwicklung ist ein Ziel der explorativen Vorstudien I und II. In Abgrenzung dazu folgen die halbstrukturierten Interviews einem Leitfaden. Jedes Interview wird mit den beiden Probanden gleichzeitig geführt, die die offene Experimentieraufgabe in Partnerarbeit bearbeiten. Pro Gruppe bewegt sich die zeitliche Länge für beide Interviewteile zusammen zwischen 12 und 14 Minuten. Zur Datenauswertung werden die Interviews transkribiert.

Ergebnisse der explorativen Vorstudie I. Die Lösung der offenen Experimentieraufgabe erfordert die Fähigkeit zur Prozessplanung, bei der Strategien zur Optimierung schnell drehender Windradprofile, wie das Sammeln und Einschätzen von Einflussgrößen bzw. -faktoren, von den Experimentatoren zu finden sind. Das kann beispielsweise durch die systematische Aufnahme und Auswertung von Messfolgen gelingen.

Entgegen den Erwartungen zeigte sich, dass typische naturwissenschaftliche Arbeitsweisen wie die Variation einer potentiellen Einflussgröße bei Konstanthalten aller anderen Faktoren (Variablenkontrollstrategie) oder die Aufnahme einer geeigneten Anzahl an Messdaten den Probanden weitgehend fremd waren. Die Planungsschritte fielen eher kurz aus und wurden schon am Ende der obligatorischen Planungsphase von einer hohen Produktorientierung dominiert: Entwürfe von gezeichneten Prototypen der Windradprofile waren für die Stichprobe charakteristisch, die von detaillierten Konstruktionsbeschreibungen begleitet wurden. Zudem stützten sich die Problemlösungen (Problemlösestrategien) auf ungeprüfte Vermutungen und Plausibilitätserklärungen (z. B. „Die echten Windräder haben alle drei Flügel, also müssen in diesem Experiment auch drei Flügel am besten sein.“). Wenn den Probanden die Planungsschritte vollständig selbst überlassen wurden, dann erkannten sie nicht alle veränderbaren Faktoren als systematisch zu untersuchende Variablen. Stattdessen konnte bei den meisten Probandenpaaren eine halbsystematische Arbeitsweise wie das gleichzeitige Variieren mehrerer Faktoren nach der eigenen Intuition identifiziert werden. Dadurch wurde vielfach eine Bewertung potentieller Einflussgrößen verhindert. Infolgedessen wurden die Windradprofile nur selten aufgrund experimenteller Messergebnisse, durch die auch neue Planungsschritte eingeleitet werden könnten, konstruiert (Kirchner & Priemer, 2007; Priemer & Kirchner, 2007a).

Um die Mängel in der Prozessplanung zu dokumentieren, wurde aus den Erhebungsdaten eine Taxonomie mit fünf Niveaustufen generiert. Differenziert wurde nach den Niveaustufen Faktoren aus einer Faktorenkombination feststellen (Niveaustufe 1), Faktoren ändern (Niveaustufe 2), Faktoren als Variablen erkennen (Niveaustufe 3), Test- und Kontrollvariablen unterscheiden (Niveaustufe 4) und Variablen bewerten (Niveaustufe 5). Um eine Zuordnung der Erhebungsdaten in die Niveaustufen vornehmen zu können, wurde in den Transkripten der Interviews gezielt nach Indikatoren gesucht, die der Tabelle 1 zu entnehmen sind. Außerdem enthält die Tabelle die Anzahl der Probandenpaare, die die Niveaustufen erreicht hatten.

Taxonomie der Planungsschritte		
Nr.	Niveaustufen (Indikatoren der Niveaustufen)	N
1	Faktoren aus einer Faktorenkombination feststellen (Beispiel: Verdrehung, Anzahl, Anstellwinkel, Länge)	9 von 9
2	Faktoren ändern (halbsystematisch: mehrere Faktoren gleichzeitig ändern)	8 von 9
3	Faktoren als Variable erkennen (Faktoren isolieren)	4 von 9
4	Test- und Kontrollvariablen unterscheiden (systematisch: Variation der TV und Konstanthalten der KV)	3 von 9
5	Variablen bewerten und Rangfolge erstellen (Beispiel: Flügellänge wichtiger als Flügelanzahl)	1 von 9
Legende: N Anzahl der Probandenpaare TV Testvariable KV Kontrollvariable		

Tab. 1: Ergebnisse der ersten empirisch-quantitativen Exploration: Taxonomie der Planungsschritte

Wie die Ergebnisse der ersten explorativen Vorstudie belegen, fällt die Anzahl der Probandenpaare von der Niveaustufe 1 zur Niveaustufe 3 ab. Erreicht wurde die Niveaustufe 3, wenn alle benannten Faktoren der Niveaustufe 1 auch als Variable identifiziert wurden. Entgegen der vorangehenden Darstellung wurde die Datenauswertung aber „probandenfreundlicher“ vorgenommen, so dass eine Eingrup-

pierung in die korrespondierende Niveaustufe bereits erfolgte, wenn von den benannten Faktoren nicht alle als Variable erkannt wurden. Trotzdem konnte die Niveausteigerung, die mit der planerischen Erstellung und der experimentellen Umsetzung von Versuchsdesigns einherging, nur von vier der 9 Probandenpaare erfüllt werden. Eine weitere Niveausteigerung war mit dem Wechsel von der Niveaustufe 4 zu der Niveaustufe 5 verbunden. Die Niveaustufe 5 galt als erreicht, wenn die Probandenpaare nur zwei Variablen unter Einhaltung der Variablenkontrollstrategie systematisch getestet und bewertet hatten. In dem geforderten Umfang war eine evidenzbasierte Datenauswertung jedoch nicht erkennbar geworden. Das hatte zur Konsequenz, dass die Planungsintentionen der Probandenpaare weder bestätigt noch verworfen werden konnten. Obwohl die Experimentierenden die Notwendigkeit von Versuchsplanung bekräftigt hatten (Priemer & Kirchner, 2007b; Priemer & Kirchner, 2007c), mussten die Planungsschritte der thematisch interessierten Stichprobe, die sich durch gute Note in Physik ausgewiesen hatte, insgesamt bemängelt werden. Aus den Ergebnissen der ersten Exploration wird gefolgert, dass Schülerinnen und Schüler wenig Erfahrung mit naturwissenschaftlichen Experimentiermethoden haben und über geringe Vorkenntnisse im Umgang mit Variablen bei offenen Experimentieraufgaben verfügen (Kirchner & Priemer, 2007).

3.2 Konsequenzen der ersten Vorstudie

Als Reaktion auf die defizitären Schülerergebnisse der explorativen Vorstudie I werden zur Förderung einer Prozessorientierung beim Experimentieren vier Lernziele formuliert, die in ein mehrstufiges bzw. mehrphasiges Vermittlungskonzept zum Umgang mit Variablen eingebunden werden. Das Konzept bildet die Basis des didaktischen Ansatzes zum Umgang mit Variablen. Die Umsetzung des Ansatzes erfolgt durch Unterricht, der in zeitlichem Abstand vor der offenen Experimentieraufgabe stattfindet. In diesem Kapitel werden die Lernziele, das Vermittlungskonzept (Inhalte, Struktur, Aufbau) und der didaktische Ansatz vorgestellt.

Lernziele. Auf Grundlage der festgestellten Niveaustufen werden für die einmalige, einführende, informierende, vorbereitende und explizite Instruktion folgende Lernziele formuliert:

1. Lernziel: Die Schüler sollen wissen, dass aus einem Set von mehreren Faktoren jeder Faktor als potentiell zu testende Variable identifiziert werden kann.
2. Lernziel: Die Schüler sollen den Unterschied zwischen Experimenten mit Kontrollansatz (nicht konfundierende Experimente) und Experimenten mit fehlerhaftem Kontrollansatz (konfundierende Experimente) kennen. Zudem sollen die Schüler wissen, dass zum Test potentieller Variablen Experimentier-

pläne mit Kontrollansatz (nicht konfundierend) geplant und erstellt werden müssen.

3. Lernziel: Die Schüler sollen wissen, dass Messfolgen von getesteten Variablen variierende Effektstärken und unterschiedliche Verläufe (linear, extremal, ohne Einfluss) haben können. Außerdem sollen die Schüler wissen, dass die Verläufe und Effektstärken von Variablen verglichen und bewertet werden müssen, wenn die Abhängigkeit der Messgröße von den getesteten Variablen gewichtet werden soll.
4. Lernziel: Die Schüler sollen erkennen, dass die Identifikation von Variablen eine Voraussetzung ist, damit Experimentierpläne mit Kontrollansatz geplant und erstellt werden können. Infolgedessen sollen die Schüler erkennen, dass Experimente mit Kontrollansatz eine Voraussetzung dafür sind, dass die Effektstärken und Verläufe von Variablen verglichen und bewertet werden können. (Zusammenhang der Lernziele 1 – 3).

Vermittlungskonzept für den Umgang mit Variablen. Die begründete Annahme, dass die Zielgruppe über wenige Vorkenntnisse im Umgang mit Variablen beim Experimentieren verfügt (vgl. Kap. 3.1), hat zur Konsequenz, dass ein in die naturwissenschaftliche Arbeitsweise einführendes Vermittlungskonzept erstellt werden muss. Deshalb wurde hinsichtlich der Auswahl und der didaktisch-methodischen Aufbereitung der Konzeptinhalte entschieden, den Umgang mit Variablen exemplarisch an einem Aufgabenbeispiel, mit einer konkreten Fragestellung und mit einer aufgabenspezifischen Auswahl von einfachen Variablen aufzuzeigen, wobei die Anzahl der ausgewählten Variablen begrenzt und im Ablauf des Vermittlungsprozesses nicht mehr verändert wird. Auf Basis dieser Kriterien werden im Folgenden drei mögliche Beispiele unter Angabe des Aufgabenthemas, der Fragestellung und der Variablenauswahl benannt:

1. Beispiel. Der Umgang mit Variablen beim Experimentieren am Beispiel der Profilkonstruktion für ein Windradmodell; Wie müssen die Flügel eines Windrades konstruiert bzw. geformt werden, damit es sich möglichst schnell dreht?; Flügelbreite, Anstellwinkel, Flügelanzahl, Flügellänge.
2. Beispiel. Der Umgang mit Variablen beim Experimentieren am Beispiel der Tragflächenkonstruktion für ein Flugzeugmodell; Wie müssen die Tragflächen eines Flugzeuges konstruiert bzw. geformt werden, damit eine möglichst große Auftriebskraft wirkt?; Spannweite, Anstellwinkel, Oberflächenbeschaffenheit, Tragflächenbreite.
3. Beispiel. Der Umgang mit Variablen beim Experimentieren am Beispiel der Konstruktion eines Wurfkörpers; Wie muss ein Wurfkörper konstruiert bzw. geformt werden, damit er möglichst weit geworfen werden kann?; Masse, Durchmesser, Oberflächenbeschaffenheit, Material des Wurfkörpers.

Die Struktur des Vermittlungskonzepts für den Umgang mit Variablen beim Experimentieren wird von der numerischen Folge der Lernziele bestimmt und durch fünf Phasen gegliedert.

In der ersten Phase wird unter Nennung des Ausgangsbeispiels ein sehr konkretes Faktorensatz mit einer begrenzten Anzahl an Faktoren vorgegeben. Das Ziel der Phase liegt in der Identifikation aller Faktoren als potentiell zu testende Variable.

In der zweiten Phase werden am Beispiel der identifizierten Variablen die Begriffe Test- und Kontrollvariable eingeführt, Experimentierpläne mit Kontrolldesign aufgezeigt und Beispiele von konfundierenden Experimentierplänen thematisiert. Im Fokus der dritten Phase steht die messtechnische Umsetzung der Experimentierpläne. Das Ziel der Phase liegt in der Demonstration der Messanordnung (Messapparatur), in der visualisierenden Inbetriebnahme der Messapparatur, in der Einführung der zu messenden Größen und in der Präsentation des Experimentiermaterials, mit dem in Voruntersuchungen nicht konfundierende Messfolgen aufgenommen wurden. Die generierten Messergebnisse werden in der vierten Phase für die Datenauswertung bereitgestellt. Es sei darauf verwiesen, dass die Durchführung von Messungen zum Zweck der Messdatenaufnahme nicht zu den Inhalten des Vermittlungskonzeptes gehört. Die Aufnahme von Messdaten ist der Bearbeitung der offenen Experimentieraufgabe vorbehalten.

Der Schwerpunkt der vierten Phase liegt auf dem Vergleich und der Bewertung von Variablen. Dazu werden drei aus eigenen Voruntersuchungen ermittelte, reale Messfolgen von stetigen und unstetigen Variablen mit variierenden Effektstärken und unterschiedlichem Verlauf (linear, extremal, ohne Einfluss) ausgewählt. Jede Messfolge wird mit maximal sechs Messdaten unter Angabe der Kontrollparameter in grafischer und tabellarischer Form präsentiert. Entscheidend ist, dass es sich bei allen präsentierten Messdaten um Maximalwerte handelt, die in den Voruntersuchungen aus mehreren Messversuchen generiert wurden. Zum Abschluss der Phase werden Kontrolldesigns mit veränderten Ausgangsparametern eingeführt. Das Ziel der Maßnahme ist es aufzuzeigen, dass Kontrolldesigns auch über die eingangs eingeführten Parameter hinaus möglich sind und dass eine Änderung der Ausgangsparameter eine Änderung der Messresultate bewirken kann. Es ist zu bedenken, dass in der vierten Phase nur eine Bewertung von Variablen vorgenommen wird und mögliche Lösungsstrategien wie die Vereinigung von Bedingungen zur Erzeugung von maximalen Werten mehrerer Variablen, die zum Bau von Windradprofilen (Beispiel 1) oder zur Konstruktion von Tragflächenprofilen (Beispiel 2) und Wurfkörpern (Beispiel 3) führen würden, nicht mehr Lernziel des vorbereitenden Vermittlungskonzeptes sind. Dadurch soll vermieden werden, dass im Vorfeld der zu bearbeitenden offenen Experimentieraufgabe mögliche Lösungen und Lösungswege transportiert werden, die die gewünschte Offenheit der Experimentierumgebung reduzieren würden. Die fünfte Phase ist eine Reflexionsphase in der der

wechselseitige Einfluss von Variablenidentifikation, Variablenkontrolle und Variablenbewertung hervorgehoben werden soll. Zum Abschluss der Phase wird die Möglichkeit zur Suche weiterer, potentiell zu testender Variablen geboten.

Neben dem beschriebenen Phasenaufbau und -ablauf ist die Wahl der Beispiele für die Konzeption des Vermittlungskonzeptes charakteristisch. Hier liegt die Besonderheit in der Fragestellung, deren Umsetzung die Produktion eigener Konstruktionen verlangen würde (Windradprofile, Tragflächenprofile, Wurfkörper). Trotzdem werden im Vermittlungskonzept für den Umgang mit Variablen nur vorgefertigte, aufgabenspezifische Experimentiermaterialien wie Sets oder Serien von Windradprofilen (Beispiel 1), von Tragflächenprofilen (Beispiel 2) oder von Wurfkörpern (Beispiel 3) eingesetzt, die zur Aufnahme nicht konfundierender Messfolgen geeignet sind. Alternativ könnte das Vermittlungskonzept auch am Beispiel eigener Konstruktionen aufbereitet werden. Jedoch muss bei der Wahl der Alternative bedacht werden, dass der Schwierigkeitsgrad der Konzeptvermittlung erhöht werden würde. Dies könnte eine zielgruppengerechte Umsetzung verhindern, weil die Probanden nach den Erkenntnissen der ersten explorativen Vorstudie nur wenig Erfahrung im Umgang mit Variablen haben (Kap. 3.1).

Didaktischer Ansatz. Aus den Ergebnissen der explorativen Vorstudie I wird gefolgert, dass die Bearbeitung der offenen Experimentieraufgabe von einem didaktischen Ansatz begleitet und gestützt werden muss. Der verfolgte Ansatz besteht aus einer Instruktionsmaßnahme, die das o. g. Vermittlungskonzept für den Umgang mit Variablen umfasst und durch die Attribute einmalig, einführend, informierend, explizit und vorbereitend gekennzeichnet werden kann. Zur Förderung prozessualer Fähigkeiten beim Experimentieren wird die Wahl eines expliziten Ansatzes als besonders wichtig erachtet, so dass experimentell-methodische Vorgehensweisen (Entwurf von Experimentierplänen und Kontrolldesigns, Erkennen von Test- und Kontrollvariablen, ...), die Offenheit der Experimentieraufgabe (Aufzeigen der Gestaltungs- und Entscheidungsmöglichkeiten, mehrere Lösungen, mehrere Lösungswege, ...), physikalisch-methodische Aspekte (Fokussierung der experimentellen Methode, Reduktion physikalischen Faktenwissens, ...) und die Ganzheitlichkeit experimenteller Abläufe (Generierung der Untersuchungsfrage, Planung, Durchführung, Auswertung und Reflexion der Experimente bzw. Experimentiertätigkeiten) als eigenständige Vermittlungsinhalte angestrebt werden können (Priemer & Kirchner, 2008).

Die weiteren Attribute, die zur Markierung des didaktischen Ansatzes benutzt werden, erklären die methodische Aufbereitung der Instruktionsmaßnahme: Im Rahmen einer Unterrichtsstunde mit einer Dauer von 45 Minuten, die im zeitlichen Abstand vor Durchführung der offenen Experimentieraufgabe stattfindet, werden die Probanden auf Basis des Vermittlungskonzepts über den Umgang mit Variablen beim Experimentieren informiert.

3.3 Explorative Vorstudie II

Das Kapitel 3.3 wird durch die drei Abschnitte Forschungsfrage, Methoden und Ergebnisse strukturiert. Folgende *Forschungsfrage* wird mit der explorativen Vorstudie II untersucht:

Welche Experimentiertätigkeiten können bei Schülern hinsichtlich des Umgangs mit Variablen bei einer offenen Experimentieraufgabe exploriert werden, wenn sie

- a) eine einmalige, einführende, informierende, explizite und vorbereitende Instruktion erhalten haben?
- b) keine vorbereitende Instruktion erhalten haben?

Methoden. *Design.* Zur Untersuchung der Fragestellungen dient die Profilkonstruktion für das Modell eines Windrades als Beispiel einer offenen Experimentieraufgabe (Kap. 2.3). Dem Design der Abbildung 3 ist zu entnehmen, dass die Probanden der Kontrollgruppe (KG) vor der offenen Experimentieraufgabe keine vorbereitende Instruktion zum Umgang mit Variablen und die Probanden der Testgruppe (TG) vor der offenen Experimentieraufgabe eine vorbereitende Instruktion zum Umgang mit Variablen erhalten. Für die Vermittlung des Variablenkonzeptes wird das erste Beispiel des Kapitels 3.2 gewählt, so dass die Testgruppe zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe keinen Kontexttransfer leisten muss. Die Experimentiertätigkeiten werden während der Bearbeitung der offenen Experimentieraufgabe durch gezielte Beobachtung und durch die Analyse des selbst erstellten schriftlichen Materials der Probandenpaare erhoben. Die Fachräume des Schülerlabors der Ruhr-Universität Bochum (AKS) dienen als Ort der Datenerhebung.

Die Probanden der Testgruppe erhalten im zeitlichen Abstand vor dem Schülerlaborbesuch unter Kontrolle der Lehrervariablen eine vorbereitende Instruktion, die in den Fachräumen der teilnehmenden Schulen stattfindet. Je nach Lerngruppe wird die zeitliche Spanne zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe, die sieben Tage nicht überschreiten sollte, flexibel festgelegt. Der zeitlich variable Abstand ist dem methodischen Anspruch der explorierenden Studie geschuldet, denn die vorbereitende Instruktion soll – nach Auskunft des schulischen Stundenplans – in der letzten Physikstunde vor dem Schülerlaborbesuch stattfinden. Die Methode minimiert den Einfluss des Fachlehrers auf die Probanden der Testgruppe, der zwischen der Vermittlung des Variablenkonzeptes und der offenen Experimentieraufgabe ausgeübt werden könnte.

Die Erprobung bzw. Exploration der vorbereitenden Instruktion wird gezielt von externen Beobachtern begleitet. Die Beobachtungsdauer beträgt 45 Minuten und entspricht der Dauer einer Unterrichtsstunde, die für die Vermittlung des Variablenkonzeptes am Beispiel der Profilkonstruktion für ein Windradmodell eingeplant ist

(Beispiel 1, Kap. 3.2). Alle Kerndaten des Untersuchungsdesigns werden mit der Abbildung 3 zusammengefasst.

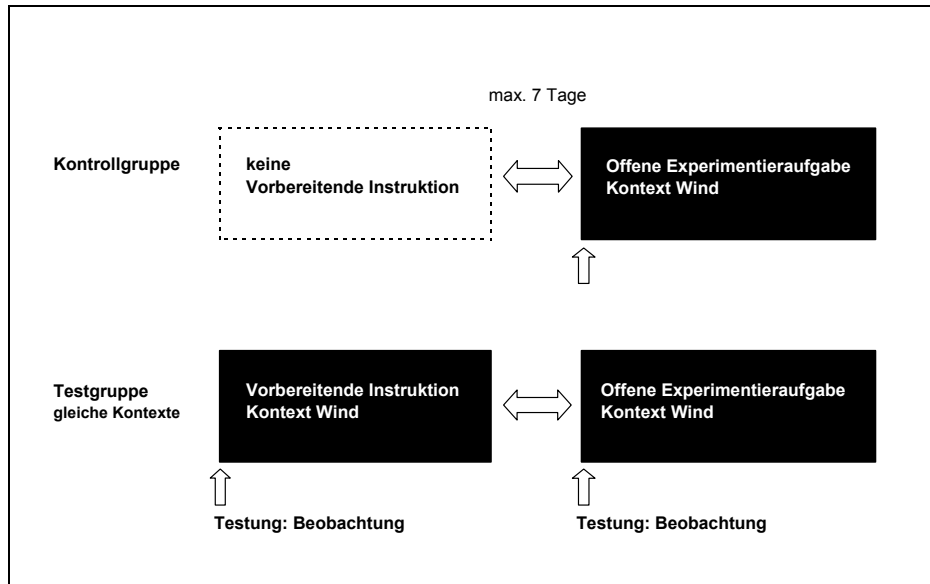


Abb. 3: Design der explorativen Vorstudie II

Stichprobe. Das Forschungsinteresse der zweiten explorativen Vorstudie liegt bei Probanden, die Physikunterricht in den Jahrgangsstufen 9 und 10 an Gymnasien und Gesamtschulen in Nordrhein-Westfalen erhalten (G 9). Da es sich um Klumpenstichproben (Schulklassen) handelt, können in dieser Erhebung klassen- und schulspezifische Effekte nicht ausgeschlossen werden. Für die Auswahl der Schulklassen ist nur die örtliche Nähe zum Schülerlaborstandort sowie die Bereitschaft zur Teilnahme an der vorbereitenden Instruktion und der offenen Experimentieraufgabe ausschlaggebend. Als Testgruppe (TG) werden drei Schulklassen mit unterschiedlicher Klassenstärke (14, 18 und 32 Schüler) rekrutiert. Dieselbe Anzahl an Schulklassen wird als Kontrollgruppe (KG) gewählt. Im Vergleich zur Testgruppe existieren geringe Unterschiede in den Klassenstärken.

Vorbereitende Instruktion und Offene Experimentieraufgabe. In diesem Abschnitt wird zum einen die Auswahl der Medien vorgestellt, die in den fünf Phasen der vorbereitenden Instruktion benutzt werden (Kap. 3.2). Zum anderen wird in diesem Abschnitt auf Differenzen verwiesen, die im Vergleich zur ersten Vorstudie mit der Auswahl und dem Einsatz der offenen Experimentieraufgabe verbunden sind.

Die ersten beiden Phasen der vorbereitenden Instruktion werden mit Unterstützung einer Powerpointpräsentation umgesetzt (Anhang). In der dritten Phase werden die Messanordnung und die Experimentiermaterialien präsentiert (fotografische Ausschnitte im Anhang). In der vierten Phase werden die zu bewertenden Messdaten mit Hilfe eines Arbeitsblattes transportiert (Anhang). In der fünften Phase wird zur

Umsetzung der Lernziele auf das Medium Powerpointpräsentation zurückgegriffen (Anhang).

In Übereinstimmung mit der ersten Vorstudie wird allen Probanden der zweiten Vorstudie – von ihrer Experimentalgruppenzugehörigkeit unabhängig – dieselbe offene Experimentieraufgabe gestellt (Kap. 2.3), die sie in altershomogenen Probandenpaaren (Partnerarbeit) bearbeiten müssen. Jedoch wird im Unterschied zur ersten Vorstudie die obligatorische Planungsphase aufgehoben, die optional von den Probandenpaaren der zweiten Vorstudie in die auf 150 Minuten limitierte Experimentierzeit mit aufgenommen werden kann. Von dieser Änderung bleibt die methodische Einbindung der offenen Experimentieraufgabe in ein Projekt unberührt (Kap. 3.1).

Messinstrumente. Mit Ausnahme der Datenerhebungsmethode Interview werden dieselben Messmethoden wie in der ersten explorativen Vorstudie eingesetzt.

Ergebnisse der explorativen Vorstudie II. Für die Erprobung des didaktischen Ansatzes ist von der Testgruppe zu berichten, dass das vorgestellte Konzept zum Umgang mit Variablen in 45 Minuten zu unterrichten war. Die Auswahlkriterien der Variablen – sehr einfache Variablen mit konkreten Messwerten und einer begrenzten, nicht veränderten Anzahl an Variablen – haben sich bewährt. Die Informationen zu den Experimentierplänen mit Kontrolldesign und den Variablenverläufen wurden von den Probanden schnell aufgenommen. Der Unterschied zwischen konfundierenden und nicht-konfundierenden Experimenten konnte ebenso vermittelt werden wie die Effektstärken von Variablen. Beides musste jedoch explizit thematisiert werden. Des Weiteren griffen die Probanden im Verlauf des Unterrichts folgende Themenfelder auf: die Abhängigkeit und Unabhängigkeit von Variablen, die Vorhersagemöglichkeit von Messergebnissen & Variablenverläufen und zur Bestimmung lokaler Extremwerte, die Differenzierung von Experimentierplänen (Kirchner & Priemer, 2008).

Im Folgenden werden die Ergebnisse des offenen Experimentierens nach der vorbereitenden Instruktion (Testgruppe TG) an den Ergebnissen des offenen Experimentierens ohne vorbereitende Instruktion (Kontrollgruppe KG) gespiegelt. Als erstes Ergebnis der Exploration ist zu nennen, dass die Planungsprozesse der Testgruppen elaborierter waren, weil im Vergleich zur Kontrollgruppe mehr Mess Tabellen zur systematischen Untersuchung von Variablen angelegt, weniger ungeprüfte Plausibilitätserklärungen geäußert und weniger Konstruktionsskizzen von Windradprofilen nach der eigenen Intuition angefertigt wurden. Als zweites Ergebnis der Exploration ist zu nennen, dass die Probanden der Testgruppe ihre Kenntnisse für den Umgang mit Variablen beim Experimentieren angewendet haben. Für die aufgestellten Lernziele des vorbereitenden Vermittlungskonzeptes konnten Indikatoren ausfindig gemacht werden. Es wurde beobachtet, dass die Trial-

& Error-Strategie für das Feststellen von Variablenabhängigkeiten, die systematischen Versuche für eine gezielte Untersuchung von Variablenverläufen, die Flügelsets für ein intensives Testen und Kontrollieren von Variablen und die experimentell erzielten hohen Spannungen für eine Bewertung der Variablen benutzt wurden. Als drittes Ergebnis der Exploration ist zu nennen, dass die Probanden offen experimentierten. Denn, es konnte nicht beobachtet werden, dass während der Experimentierphase die Profile des bereitgestellten, aufgabenspezifischen Experimentiermaterials Flügelsets nachgebaut oder kopiert wurden. Dies wurde als Hinweis dafür gewertet, dass die Probanden zur Konstruktion der Windradprofile nach eigenen Lösungen und Lösungswegen gesucht haben. Zudem lassen nicht vorhandene Aussteigereffekte darauf schließen, dass sich die Probanden während der Experimentierphase weder über- noch unterfordert gefühlt haben (Kirchner & Priemer, 2008). Unter dem Begriff Aussteigereffekte werden Effekte subsummiert, die zum Abbruch bzw. zur vorzeitigen Beendigung der Experimentiertätigkeiten führen, so dass die Probanden auf eine Lösung der angebotenen offenen Experimentieraufgabe verzichten. Die positiven Ergebnisse, die beim Experimentieren der Testgruppe festgestellt wurden, werden auf die Wirkung des didaktischen Ansatzes zurückgeführt.

3.4 Ergänzungen und kritische Anmerkungen

Die Ergänzungen und kritischen Anmerkungen beziehen sich auf die Erhebungsdaten, die Erhebungsmethoden und die Ergebnisse beider explorativen Vorstudien. Vor allem die kritischen Anmerkungen dienen dem Ziel, die Forschungsfragen und -methoden der Hauptstudie zu generieren, zu spezifizieren und zu differenzieren.

Anders als in den explorierenden Vorstudien geschehen, müssen die beobachtbaren Experimentiertätigkeiten für eine wirksamkeitsüberprüfende Hauptstudie in zwei Phasen eingeteilt werden. Zur ersten Experimentierphase werden die Experimentierprozesse gehören, die die Probanden im Umgang mit Variablen zeigen, wenn sie zur Bearbeitung der offenen Experimentieraufgabe das aufgabenspezifische Experimentiermaterial Flügelsets und die angebotenen Hilfsmittel verwenden. Die Experimentierphase wird mit der Bezeichnung „Profilkonstruktion vorbereiten“ beschrieben. Ferner werden die planerischen Tätigkeiten der Probanden zu den Inhalten der Phase gezählt. Zur zweiten Experimentierphase werden die Experimentierprozesse und -tätigkeiten gehören, die die Probanden im Umgang mit Variablen zeigen, wenn sie zur Lösung der offenen Experimentieraufgabe eigene Profile konstruieren. Mit der Bezeichnung „Profil konstruieren“ wird die Phase angemessen beschrieben. Die

Einteilung in die Experimentierphasen wird notwendig, wenn der Einfluss der vorbereitenden Instruktion auf die beobachtbaren Experimentiertätigkeiten differenzierter herausgestellt werden soll. Denn dem Konzept zum Umgang mit Variablen ist zu entnehmen, dass für die Vermittlung von kontrollierten Experimentierplänen und der Bewertung von Variablen nur vorgefertigtes, aufgabenspezifisches Experimentiermaterial wie die Sets von Windradprofilen (Beispiel 1), von Tragflächenprofilen (Beispiel 2) oder von Wurfkörpern (Beispiel 3) verwendet wird (Kap. 3.2). Diese Einschränkung ist gewollt, damit die Fähigkeiten zum Transfer des Variablenkonzeptes von der vorbereitenden Instruktion auf die offene Experimentieraufgabe unter dem Einfluss veränderter Experimentierbedingungen gemessen werden können (vgl. Kap. 4.4.4). Die Änderung der Experimentierbedingungen ist durch die wechselseitigen, probandenpaarspezifischen Übergänge der Experimentierphasen Profilkonstruktion vorbereiten (Experimentierphase 1) und Profil konstruieren (Experimentierphase 2) gegeben.

Des Weiteren muss angemerkt werden, dass im Rahmen der explorierenden Studien eine Kategorisierung der festgestellten Niveaustufen gelungen ist. Um weiterreichende Analysen hinsichtlich des Umgangs mit Variablen beim Experimentieren interpretieren zu können, werden zusätzliche Kategorisierungen der beobachtbaren Experimentierprozesse (z. B. Aufnahme von Messdaten) und der verbalen Argumentationen (z. B. Vergleiche und Bewertungen von Messdaten) notwendig. Beispielsweise werden die experimentierenden Probanden durch eine Kategorisierung ihrer verbalen Vergleiche und Bewertungen unterscheidbarer, wenn eine Differenzierung zwischen dem Vergleich von Einzelmessungen und dem Vergleich von Messfolgen geschaffen wird. Um im Vorfeld der Hauptstudie zu weiteren systematischen Hinweisen zu gelangen, werden im Anschluss der beiden Explorationen weitere Lerngruppen beobachtet. Die Beobachtungsergebnisse führen zur Konzeption des Messinstruments Beobachtungsbogen (Kap. 4.2.3), das für die Datenerhebung der Hauptstudie benutzt wird. Nach dem Design der zweiten explorativen Vorstudie werden zu diesem Zweck bis zur Erstellung der finalen Version des Messinstruments, die Experimentiertätigkeiten von Probandenpaaren aus 22 verschiedenen Lerngruppen gezielt beobachtet. Zusammen sind das mehr als 300 Teilnehmer, die die offene Experimentieraufgabe bearbeiten. Da mit dem Messinstrument auch stark divergierende Experimentierprozesse differenziert erhoben und ausgewertet werden sollen, werden für den Zeitraum, in dem das Messinstrument konzipiert wird, bewusst stark heterogene Lerngruppen rekrutiert. Zu den ausgewählten Lerngruppen gehören Klassen und Kurse der Schulformen Gymnasium und Gesamtschule (Jahrgangsstufen 5 – 13) sowie Klassen und Kurse der Schulformen Haupt- und Realschule (Jahrgangsstufe 9 und 10) aus Nordrhein-Westfalen. Allerdings wurden im Hinblick auf die Forschungsinteressen der Hauptstudie bevorzugt Gymnasialbesucher der 10. Jahrgangsstufe (G 9) beobachtet.

Eine weitere kritische Anmerkung richtet sich an die Datenerhebungsmethode Interview, die im Rahmen explorierender Vorstudien als zulässig eingestuft wird. Jedoch wird mit der Wahl des Messinstruments Beobachtungsbogen für die Hauptstudie eine Methode favorisiert, die erstens die Reaktivität des Datenaufnehmers reduziert und zweitens, bedingt durch den Einsatz trainierter Beobachter, nicht mehr den potentiell verzerrten Antworten einer Befragung unterliegt.

Die ebenfalls als Alternative zum Interview erprobte, nicht reaktive Datenerhebungsmethode des eigenständigen, schriftlichen Protokollierens, von der bisher noch nicht berichtet wurde, verlief im Vergleich zur gezielten Beobachtung der Experimentierprozesse wenig erfolgreich, denn: Die von den Probanden selbst verfassten Protokolle wiesen – obwohl ein unterstützender Leitfaden angeboten wurde – massive qualitative und quantitative Defizite auf, so dass für den Umgang mit Variablen beim Experimentieren eine Analyse und Interpretation der Erhebungsdaten nicht möglich war. Die erprobte Methode schlug vermutlich fehl, weil die Probanden mit den hohen Anforderungen des selbständigen Reflektierens ihrer Experimentiertätigkeiten und -strategien wie Variablen erkennen, Variablen testen, Variablen kontrollieren und Variablen bewerten überfordert waren. Die Verwertbarkeit der Protokolle kann vermutlich mit einem adressatengerechten Methodentraining gesteigert werden. Jedoch muss auch beachtet werden, dass der nicht einfach zu kontrollierende, reaktive Einfluss eines Probandentrainings, der Ergebnisverzerrungen bewirken kann, nicht unterschätzt werden darf. Aus diesem Grund wurde die Möglichkeit des selbständigen Protokollierens von Experimentierprozessen durch Probanden als geeignete Datenerhebungsmethode für die Hauptstudie verworfen. Ausgenommen hiervon sind die gezeichneten Entwürfe von Prototypen der Windradprofile (Kap. 3.1) und die durch Tabellen und grafische Abbildungen schriftlich fixierten Messergebnisse der Probanden.

Abschließend muss zur Erprobung des didaktischen Ansatzes kritisch angemerkt werden, dass für die vorbereitende Instruktion und für die offene Experimentieraufgabe identische Kontexte verwendet wurden (beide: Profilkonstruktion für das Modell eines Windrades). Deshalb konnten positive Effekte beim offenen Experimentieren erwartet werden.

3.5 Ausblick Hauptstudie

Ein Untersuchungsinteresse der Hauptstudie liegt in der Analyse von Fähigkeiten, die Probanden beim Transfer eines Variablenkonzeptes von einer vorbereitenden Instruktion auf eine offene Experimentieraufgabe unter der Bedingung zeigen, dass zwischen vorbereitender Instruktion und offener Experimentieraufgabe ein Kontextwechsel stattfindet (Kontexttransfer). Um die Graduierung der Kontext-

unterschiede anzuzeigen, werden die Begriffe Reproduktion (gleiche Kontexte), Reorganisation (ähnliche Kontexte) und Transfer (verschiedene Kontexte) verwendet. Zur Erzeugung der Kontextunterschiede werden die Ausprägungen des vermittelnden Variablenkonzeptes für die Testgruppen TG 1 (Reproduktion), TG 2 (Reorganisation) und TG 3 (Transfer) variiert, wohingegen von allen Experimentiergruppen dieselbe offene Experimentieraufgabe zu bearbeiten ist. Als Beispiel der offenen Experimentieraufgabe wird die Profilkonstruktion für das Modell eines Windrades gewählt, so dass bei den gegebenen Versuchsbedingungen eine möglichst hohe Spannung erzeugt wird (Kap. 2.3 und Kap. 3.1).

Eine Herausforderung des Forschungsvorhabens ist die Konzeption eines für alle Testgruppen vergleichbaren, vorbereitenden Variablenkonzeptes, dass den Kontexttransferleistungen genügt. Das Ziel liegt zum einen in der Darstellung der Zuordnungskriterien der Transferleistungen und zum anderen in der Darstellung der Vergleichbarkeitskriterien der vorbereitenden Instruktion. Die Zuordnungen der Transferleistungen werden vollzogen, indem die Kontextunterschiede zwischen den Ausprägungen der Instruktionen und der offenen Experimentieraufgabe bewertet werden. Dazu müssen unterschiedliche Lernvoraussetzungen der Zielgruppe (Gymnasialbesucher nach Abschluss der Jahrgangsstufe 10) aufgrund der Erkenntnisse, die aus den Explorationen zum Umgang mit Variablen generiert wurden (Kap. 3.1) und aufgrund der Auswahl der Aufgabenbeispiele, die mit den curricularen Vorgaben abgestimmt werden, nicht gewichtet werden (s. u.).

Um bei variierenden Kontexten die Vergleichbarkeit einer vorbereitenden Instruktion zu garantieren, wird ein selbst erstelltes Vermittlungskonzept zum Umgang mit Variablen benutzt (Kap. 3.2), dessen Inhalte, Struktur und Phasenablauf auf alle Konzeptausprägungen übertragen wird. Zudem wird die Vergleichbarkeit der vorbereitenden Instruktion durch die Auswahl der phasenbegleitenden Medien erhöht, indem alle Ausprägungen der Medienauswahl des Kapitels 3.3 folgen. Die Kontexte der Ausprägungen werden durch die Wahl der Aufgabenbeispiele und Fragestellungen bestimmt, mit denen der Umgang mit Variablen beim Experimentieren exemplarisch vermittelt wird. Für die Testgruppe 1 (Reproduktion) wird die Fragestellung „Wie müssen die Flügel eines Windrades konstruiert bzw. geformt werden, damit es sich möglichst schnell dreht?“ (Beispiel 1, Kap. 3.2), für die Testgruppe 2 (Reorganisation) wird die Fragestellung „Wie müssen die Tragflächen eines Flugzeuges konstruiert bzw. geformt werden, damit eine möglichst große Auftriebskraft wirkt?“ (Beispiel 2, Kap. 3.2) und für die Testgruppe 3 (Transfer) wird die Fragestellung „Wie muss ein Wurfkörper konstruiert bzw. geformt werden, damit er möglichst weit geworfen werden kann?“ (Beispiel 3, Kap. 3.2) gewählt.

Mit der Wahl der Fragestellungen ist die Hürde verbunden, dass für alle Aufgabenbeispiele reale Messfolgen von stetigen und unstetigen Variablen mit unter-

schiedlichen Verläufen (linear, extremal, ohne Einfluss) gefunden werden müssen, so dass die Vergleichbarkeit der Ausprägungen des vermittelten Variablenkonzeptes auch in dieser Hinsicht erfüllt wird. Der Aspekt wird in den folgenden Zuordnungen der Transferleistungen aufgegriffen. Darüber hinaus können die Folien der Powerpointpräsentationen und Arbeitsblätter, die die vorbereitenden Instruktionen unterstützend begleiten, im Anhang eingesehen werden. Die Zuordnungen der Transferleistungen werden durch eine Gegenüberstellung mit der offenen Experimentieraufgabe begründet.

Transferleistung Reproduktion (Testgruppe 1). Um der Fragestellung der vorbereitenden Instruktion mit der Transferleistung Reproduktion nachzugehen, werden den Probanden drei Sets von vorgefertigten Windradprofilen präsentiert (Flügelsets), mit denen systematisch die Testvariablen Flügelbreite, Anstellwinkel und Flügelanzahl untersucht werden. Pro Messfolge werden maximal sechs Messwerte abgebildet. Mit dem ersten Set werden die maximalen Drehzahlen der Testvariablen Flügelbreite als Beispiel für eine stetig veränderbare Variable mit linearem Verlauf, mit dem zweiten Set werden die maximalen Drehzahlen der Testvariablen Anstellwinkel als Beispiel für eine stetig veränderbare Variable mit extremalen Verlauf und mit dem dritten Set werden die maximalen Drehzahlen der Testvariablen Flügelanzahl als Beispiel für eine nicht stetig veränderbare Variable gegeben, die in einem begrenzten Intervall keinen Einfluss auf die maximale Drehzahl ausübt. Die Variation der Testvariablen und deren Ausprägungen bewirken maximale Drehzahlen von 1000 U/ min bis 2460 U/ min. Als Messanordnung (Messeinrichtung) zur Aufnahme der maximalen Drehzahlen dient das Modell einer Windradanlage (fotografische Abbildung im Anhang).

Die Zuordnung der Transferleistung ist begründet, weil die Kontexte und Fragestellungen von vorbereitender Instruktion und offener Experimentieraufgabe identisch sind. Das in den Fragestellungen verankerte physikalische Prinzip des dynamischen Auftriebs zählt nicht zu den curricularen Vorgaben des Physikunterrichts. Das aufgabenspezifische Experimentiermaterial Flügelset wird sowohl in der Vorbereitung als auch in der offenen Experimentieraufgabe (Experimentierphase 1) benutzt. Zur Aufnahme der Messwerte werden dieselben Messanordnungen verwendet. Die vorbereitende Instruktion unterscheidet sich von der offenen Experimentieraufgabe u. a. durch die Anzahl der eingesetzten bzw. bereitgestellten Flügelsets und durch die Messgröße, mit der die maximalen Umdrehungsgeschwindigkeiten der Windradprofile bestimmt werden. In der Vorbereitung wird die Drehzahl und in der offenen Experimentieraufgabe wird die Spannung als Messgröße der maximalen Umdrehungsgeschwindigkeit benutzt. In der Vorbereitung wird die Messgröße Spannung vermieden, weil der Schwerpunkt auf der Vermittlung naturwissenschaftlicher Arbeitsweisen und nicht auf der Vermittlung fachlicher Inhalte liegt. Die theoretischen Basisinformationen wie Generatorprinzip

oder dynamischer Auftrieb werden direkt vor der Aufgabenstellung der offenen Experimentieraufgabe mit einem Lehrerexperimentalvortrag transportiert (vgl. Projekteinführungsphase Kap. 3.1).

Transferleistung Reorganisation (Testgruppe 2). Um der Fragestellung der vorbereitenden Instruktion mit der Transferleistung Reorganisation nachzugehen, werden den Probanden drei Sets von vorgefertigten Tragflächenprofilen präsentiert (Tragflächensets), mit denen systematisch die Testvariablen Spannweite, Anstellwinkel und Oberflächenbeschaffenheit untersucht werden. Pro Messfolge werden maximal sechs Messwerte abgebildet. Mit dem ersten Set werden die maximalen Auftriebskräfte der Testvariablen Spannweite als Beispiel für eine stetig veränderbare Variable mit linearem Verlauf, mit dem zweiten Set werden die maximalen Auftriebskräfte der Testvariablen Anstellwinkel als Beispiel für eine stetig veränderbare Variable mit extremalem Verlauf und mit dem dritten Set werden die maximalen Auftriebskräfte der Testvariablen Oberflächenbeschaffenheit (Anzahl der Querrillen) als Beispiel für eine nicht stetig veränderbare Variable gegeben, die in einem begrenzten Intervall keinen Einfluss auf die maximale Auftriebskraft ausübt. Die Variation der Testvariablen und deren Ausprägungen bewirken maximale Auftriebskräfte von 3,7 cN bis 15,8 cN. Als Messanordnung (Messeinrichtung) zur Aufnahme der maximalen Massenverluste (Messgröße) dient eine Tragflächenhalterung (fotografische Abbildung im Anhang).

Die Zuordnung der Transferleistung ist begründet, weil mit der vorbereitenden Instruktion und der offenen Experimentieraufgabe unterschiedliche Fragestellungen verfolgt werden, die demselben Kontext angehören. Das in beiden Fragestellungen verankerte physikalische Prinzip des dynamischen Auftriebs zählt nicht zu den curricularen Vorgaben des Physikunterrichts. Das aufgabenspezifische Experimentiermaterial Tragflächensets wird nur in der Vorbereitung verwendet, wohingegen in der offenen Experimentieraufgabe die Flügelsets von Windradprofilen zur Verfügung gestellt werden (Experimentierphase 1). Zur Aufnahme der Messwerte werden sowohl die Tragflächenprofile als auch die Windradprofile mit einem Gebläse umströmt. Obwohl für beide Objekte (Tragflächen und Windradprofile) dieselben Testvariablen untersucht werden können, müssen die Unterschiede der Messgrößen beachtet werden.

Transferleistung Transfer (Testgruppe 3). Um der Fragestellung der vorbereitenden Instruktion mit der Transferleistung Transfer nachzugehen, werden den Probanden drei Sets von vorgefertigten Wurfkörpern präsentiert (Wurfkörpersets), mit denen systematisch die Testvariablen Masse, Oberflächenbeschaffenheit und Durchmesser untersucht werden. Pro Messfolge werden maximal sechs Messwerte abgebildet. Mit dem ersten Set werden die maximalen Wurfweiten der Testvariablen Masse als Beispiel für eine stetig veränderbare Variable mit linearem Verlauf (intervall-

begrenzt), mit dem zweiten Set werden die maximalen Wurfweiten der Testvariablen Oberflächenbeschaffenheit (Anzahl der Löcher) als Beispiel für eine nicht stetig veränderbare Variable mit extremalem Verlauf und mit dem dritten Set werden die maximalen Wurfweiten der Testvariablen Durchmesser als Beispiel für eine stetig veränderbare Variable gegeben, die in einem begrenzten Intervall keinen Einfluss auf die maximale Wurfweite ausübt. Die Variation der Testvariablen und deren Ausprägungen bewirken maximale Wurfweiten von 624 cm bis 1163 cm. Als Messanordnung (Messeinrichtung) zur Aufnahme der maximalen Wurfweiten dient ein Katapult (fotografische Abbildung im Anhang).

Die Zuordnung der Transferleistung ist begründet, weil mit der vorbereitenden Instruktion und der offenen Experimentieraufgabe unterschiedliche Fragestellungen verfolgt werden, die verschiedenen Kontexten angehören. Während das physikalische Prinzip des dynamischen Auftriebs nicht zu den curricularen Vorgaben des Physikunterrichts zählt, können Wurfbewegungen optional in der gymnasialen Oberstufe als Teilgebiet der Kinematik des Massenpunktes thematisiert werden.

Für den Fall, dass die Kontrolle des Vorwissens vernachlässigt werden würde, wird das Thema in der vorbereitenden Instruktion mit gravierenden Unterschieden zu den curricularen Vorgaben des Physikunterrichts behandelt. Anders als im regulären Unterricht, werden die wurfweitenbeeinflussenden Faktoren nicht in der Variation von Versuchsbedingungen wie Abwurfwinkel und Abwurfgeschwindigkeit, sondern in der Variation von Testvariablen wie Durchmesser und Oberflächenbeschaffenheit des Wurfkörpers unter konstanten Versuchsbedingungen gesucht. Bei der praktizierten Variante ist jedoch zu bedenken, dass aerodynamische Effekte wie der Magnus-Effekt die Wurfweite des Körpers beeinflussen können. Den Probanden ist dies in der Regel nicht bekannt, weil aerodynamische Effekte wie der Magnus-Effekt nicht zu den curricularen Vorgaben des Physikunterrichts gehören.

Die Zuordnung der Transferleistung ist gerechtfertigt, weil die Probanden zu wenige Vorkenntnisse haben, so dass der Zusammenhang zwischen der Messgröße „maximale Wurfweite eines Wurfkörpers“ (Vorbereitung) und der Messgröße „maximale Spannung eines Windradprofils“ (offene Experimentieraufgabe) nur schwer hergestellt werden kann. Das aufgabenspezifische Experimentiermaterial Wurfkörpersets wird nur in der Vorbereitung verwendet, wohingegen in der offenen Experimentieraufgabe die Flügelsets von Windradprofilen zur Verfügung gestellt werden (Experimentierphase 1). Zur Aufnahme der Messwerte werden die Windradprofile der Flügelsets mit einem Gebläse umströmt. Davon unterscheiden sich die aerodynamischen Auftriebseffekte, die durch die Relativbewegung von Wurfkörpern der Wurfkörpersets (bewegend) und der Luft („ruhend“) erzeugt werden.

Die drei Ausprägungen der vorbereitenden Instruktion wurden mit 72 Schülern der 10. Jahrgangsstufe eines Düsseldorfer Gymnasiums erprobt. Das Ziel der Explo-

ration ist, auch auf der Basis von Schülerreaktionen einen Abgleich der Instruktionen herzustellen, die durch den vorbereitenden Unterricht zum Umgang mit Variablen beim Experimentieren provoziert werden. Das Ziel der Exploration erfordert nicht, dass die Probanden die offene Experimentieraufgabe tatsächlich bearbeiten. Da die drei Ausprägungen des Variablenkonzeptes unter ähnlichen Bedingungen zur Hauptstudie unterrichtet werden sollen, werden aus der Stichprobe drei Experimentalgruppen gebildet. Die Experimentalgruppen werden unter Auflösung von Klasseneffekten nach ihren fachlichen Leistungen in Physik zusammengesetzt. Zum Erhebungszeitpunkt ist bereits bekannt, dass die ausgewählten Probanden die Sekundarstufe I am Gymnasium (G 9) erfolgreich abschließen werden. Jede Variante der drei vorbereitenden Instruktionen wird einmalig getestet und von demselben externen Beobachter begleitet.

Ein Ergebnis der Beobachtung war, dass die Abhängigkeit und Unabhängigkeit von Variablen bevorzugt von Probanden aufgegriffen wurde, die die vorbereitende Instruktion am Beispiel der Konstruktion eines Wurfkörpers erhalten haben. Als ein Beispiel sei genannt, dass die Probanden geäußert haben, die Testung der Variablen Oberflächenbeschaffenheit (Anzahl der Löcher) könne eine Änderung der Masse des Wurfkörpers bewirken, die nach den Ergebnissen der Voruntersuchungen unterhalb von 0,3 % liegt. Die Reaktion der Schüler hat hinsichtlich des Abgleichs der Instruktionen für die Hauptstudie zur Konsequenz, dass die Thematik Abhängigkeit von Variablen in den vorbereitenden Instruktionen am Beispiel der Profilkonstruktion für ein Windradmodell (TG 1, Reproduktion) und am Beispiel der Tragflächenkonstruktion für ein Flugzeugmodell (TG 2, Reorganisation) stärker problematisiert werden muss als am Beispiel der Konstruktion eines Wurfkörpers (TG 3, Transfer).

4 Hauptstudie

4.1 Forschungsfragen

Mit der Hauptstudie werden folgende Forschungsfragen untersucht:

1. Welche Fähigkeiten im Umgang mit Variablen zeigen Schüler am Ende der Sekundarstufe I (G 9) bei einer offenen Experimentieraufgabe mit einer einmaligen, einführenden, informierenden, expliziten, vorbereitenden Instruktion und ohne durchlaufene Instruktion?
2. Wie inhaltlich nah – im Sinne von Reproduktion, Reorganisation und Transfer – muss die vorbereitende Instruktion an der offenen Experimentieraufgabe liegen, damit Schüler am Ende der Sekundarstufe I bei der offenen Experimentieraufgabe elaborierte Fähigkeiten im Umgang mit Variablen zeigen?
3. Welche Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen sind bei Schülern am Ende der Sekundarstufe I (G 9) festzustellen, die
 - a) eine einmalige, einführende, informierende, explizite, vorbereitende Instruktion erhalten, die inhaltlich verschieden nah – im Sinne von Reproduktion, Reorganisation und Transfer – an einer offenen Experimentieraufgabe liegt und die diese Aufgabe anschließend bearbeiten?
 - b) eine offene Experimentieraufgabe bearbeiten, ohne mit einer Instruktion vorbereitet zu werden?
4. Wie nachhaltig sind die durch die verschiedenen Instruktionen (vgl. Forschungsfragen 3 a/ b) erworbenen allgemeinen Fähigkeiten im Umgang mit Variablen bei Schülern am Ende der Sekundarstufe I?

Zu den Forschungsfragen sind folgende Anmerkungen notwendig:

- a) Die Hauptinteressen der Studie liegen bei den ersten beiden Forschungsfragen. Mit den letzten beiden Forschungsfragen soll exploriert werden, ob die Untersuchungseinheit aus vorbereitender Instruktion und offener Experimentieraufgabe über die beobachtbaren Experimentierhandlungen hinaus eine Wirkung

zeigt, die mit dem schriftlichen Messinstrument Fach- & NAW-Test erhoben werden kann.

- b) In den Forschungsfragen wird der Begriff Transferleistung benutzt, um die Fähigkeiten der Probanden zum Transfer eines Variablenkonzeptes zu dokumentieren, der unter der Bedingung eines Kontextwechsels zwischen vorbereitender Instruktion und offener Experimentieraufgabe geleistet werden muss (Kontexttransfer). Um die Graduierung der Kontextunterschiede zwischen vorbereitender Instruktion und offener Experimentieraufgabe anzuzeigen, werden die Begriffe Reproduktion (gleiche Kontexte), Reorganisation (ähnliche Kontexte) und Transfer (verschiedene Kontexte) verwendet.
- c) In den letzten beiden Forschungsfragen wird die Bezeichnung allgemeine Fähigkeiten im Umgang mit Variablen benutzt, um die Fähigkeiten der Probanden zu dokumentieren, die durch die Items des Messinstruments Fach- & NAW-Test erfasst werden. Das Messinstrument basiert auf einer selbst zusammengestellten Kollektion von Items für die Dimension Fachwissen (Fachwissens-Test) und für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW-Test). Die allgemeinen Fähigkeiten im Umgang mit Variablen werden mit der Dimension Naturwissenschaftliche Arbeitsweise erhoben (vgl. Kap. 4.2.2).

4.2 Methoden

Aufgrund der unbefriedigenden Forschungslage zum Umgang mit Variablen bei offenen Experimentieraufgaben in Physik, wird mit der vorliegenden Studie ein explorativer Ansatz verfolgt. In der fachdidaktischen Forschung mangelt es an Ergebnissen zur Klassifikation von prozessbezogenen Fähigkeiten beim Experimentieren (Niveaustufen des Experimentierens), die auf direkten Beobachtungen und nicht auf deren Ersatzmaßen fußen. In dessen Folge mangelt es an Forschungsergebnissen zu Angeboten an didaktischen Ansätzen, Instruktionen und Fördermaßnahmen, die zu einer Verbesserung bzw. Steigerung der naturwissenschaftlichen Arbeitsweise Umgang mit Variablen beim Experimentieren führen. Weiterhin kann eine ernüchternde Datenlage zur Forschung der Transferleistungen beim Experimentieren (Kontexttransfer) festgestellt werden.

Mit der Wahl eines explorativen Untersuchungstyps möchte diese Studie für die genannten Themenfelder mit einer evidenzbasierten Generierung von Forschungshypothesen überzeugen, indem direkt beobachtete Experimentierprozesse analysiert, bewertet und in Experimentierniveaustufen überführt werden. Hierbei wird als wichtig erachtet, dass die Messdaten im Gegensatz zu Pre-Post oder ähnlichen

Designs in situ des Experimentierens von trainierten Beobachtern erhoben werden. Ziel der Studie ist es nicht, die Hypothesen zu überprüfen.

4.2.1 Design, Stichprobe, Teilstichprobe, Variablen

Design. Das Design zu den ersten beiden Forschungsfragen wird durch eine Komposition aus „Vorbereitender Instruktion“ und „Offener Experimentieraufgabe“ wiedergegeben. Die vorbereitende Instruktion findet in den Fachräumen der teilnehmenden Physikkurse unter Kontrolle des zeitlichen Umfangs (eine Schulstunde mit 45 Minuten Unterrichtszeit) und unter Kontrolle der Lehrervariablen statt. Die offene Experimentieraufgabe wird von allen Probanden im Schülerlabor der Ruhr-Universität Bochum (Alfried-Krupp Schülerlabor, AKS) bearbeitet, welches als Ort der Datenerhebung mit Kontrolle der Experimentierbedingungen bzw. mit Kontrolle der (Schüler-) Laborvariablen dient. Im AKS werden die Fähigkeiten im Umgang mit Variablen beim Experimentieren unter Einsatz des Messinstruments Beobachtungsbogen über eine maximale Experimentierzeit von 150 Minuten erhoben (vgl. Kap. 4.2.3, Kap. 4.4). Zwischen beiden Komponenten liegt ein zeitlicher Abstand, der für die Test- und Kontrollgruppen zwischen drei und sieben Tagen variiert. Die variierenden zeitlichen Spannen sind dem Ziel geschuldet, dass zwischen der vorbereitenden Instruktion in der Schule und der Bearbeitung der offenen Experimentieraufgabe im Schülerlabor nur ein Minimum an Physikunterricht stattfinden soll, der vom jeweiligen Kurslehrer erteilt wird. Nach den schulinternen Stundenplänen findet für die Mehrheit der beteiligten Physikkurse zwischen der vorbereitenden Instruktion und dem Schülerlaborbesuch kein Physikunterricht statt.

Um die Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen nach den Forschungsfragen 3 und 4 zu erheben, wird dem Design aus vorbereitender Instruktion und offener Experimentieraufgabe ein Pre-Post-Follow-Up Design hinzugefügt. Zu den drei Messzeitpunkten wird das Messinstrument Fach- & NAW-Test wiederholt eingesetzt (vgl. Kap. 4.2.2, Kap. 4.3). Die erste Datenerhebung findet unmittelbar vor der vorbereitenden Instruktion in der Schule (Vortest, Messzeitpunkt 1), die zweite Datenerhebung unmittelbar nach der Bearbeitung der offenen Experimentieraufgabe im Schülerlabor (Nachtest, Messzeitpunkt 2) und die dritte Datenerhebung drei Monate nach der Bearbeitung der offenen Experimentieraufgabe in der Schule statt (Verzögerter Nachtest, Messzeitpunkt 3). Es sei darauf verwiesen, dass für eine hypothesenüberprüfende Studie neben der Kontrollgruppe, die die offene Experimentieraufgabe ohne vorbereitende Instruktion bearbeitet, mindestens eine weitere Kontrollgruppe notwendig wäre, die nur eine einmalige, einführende, informierende, explizite und vorbereitende Instruktion erhält und an der offenen Experimentieraufgabe nicht partizipiert. Da es sich bei der

vorliegenden Studie um einen hypothesengenerierenden Untersuchungstyp handelt, wurde auf eine doppelte Absicherung durch eine zweite Kontrollgruppe verzichtet, so dass die Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen, die auf die vorbereitende Instruktion zurückzuführen sind, aus den Vergleichen der Kontrollgruppe mit den Testgruppen generiert werden.

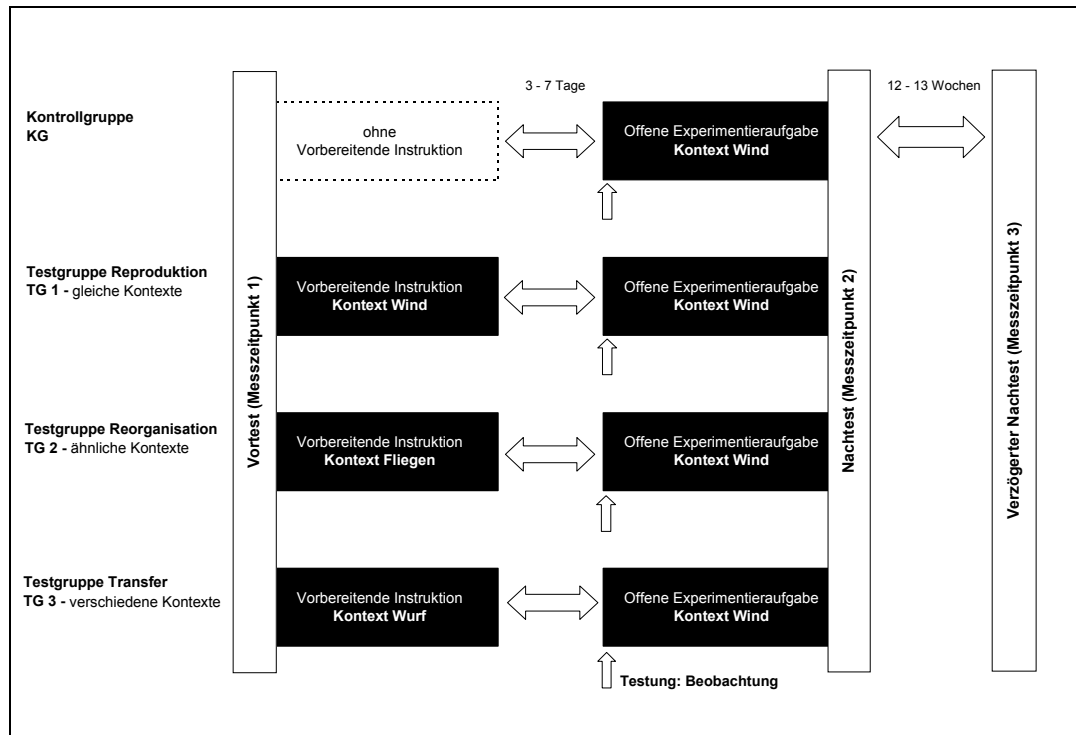


Abb. 4: Design der Hauptstudie

Die Datenerhebung der Hauptstudie erfolgte im Schuljahr 2008/ 2009. Die Erhebungseinheit aus Vortest, vorbereitender Instruktion, Bearbeitung der offenen Experimentieraufgabe und Nachttest wurde für alle acht Experimentalgruppen (Test- und Kontrollgruppen) in einem vierwöchigen Zeitabschnitt zwischen den Sommer- und Herbstferien der Schule vom 25. August 2008 bis zum 26. September 2008 durchgeführt. Der Verzögerte Nachtest wurde von allen Experimentalgruppen kurz vor den (Schul-) Weihnachtsferien in einem zweiwöchigen Zeitabschnitt vom 4. Dezember 2008 bis zum 17. Dezember 2008 erhoben. Alle Kerndaten des Untersuchungsdesigns werden mit der Abbildung 4 zusammengefasst. Der Abbildung ist nicht zu entnehmen, dass für die Untersuchung ein quasi-experimentelles Design gewählt wurde, bei dem die Experimentalgruppen durch Randomisierung ermittelt wurden. Zudem muss der Abbildung hinzugefügt werden, dass es sich bei den Test- und Kontrollgruppen nicht mehr um die natürlichen Gruppen der Stichprobe (Physikkurse), sondern um eine Teilstichprobe handelt (vgl. *Stichprobe* und *Teilstichprobe*). Mit der kriteriengeleiteten Ziehung der Teilstichprobe wird eine

hohe interne Validität anvisiert, die für eine hypothesengenerierende Studie als notwendig und wichtig erachtet wird.

Stichprobe. Das Forschungsinteresse liegt bei Probanden, die Physikunterricht in der Sekundarstufe I (G 9) an Gymnasien in Nordrhein-Westfalen erhalten haben und die Jahrgangsstufe 10 mit der Qualifikation für die gymnasiale Oberstufe beendet haben. Um die Probendenauswahlkriterien zu garantieren, werden aus der Population der Gymnasialschüler in NRW Probanden an der Schnittstelle zur Sekundarstufe II rekrutiert, die zum Erhebungszeitpunkt bereits einen Physikgrundkurs in der Jahrgangsstufe 11 besuchen. Die Besonderheit dieser Probanden liegt zum einen in ihrer Zugehörigkeit zu Lerngruppen, die mit Beginn der gymnasialen Oberstufe neu zusammengestellt werden. Dies bietet den Vorteil, dass die Kurse noch nicht von klassen- bzw. klumpenspezifischen Effekten geprägt sind, die eine Ergebnisverzerrung hervorrufen könnten. Zum anderen liegt die Besonderheit dieser Probanden in ihrer Wahlentscheidung, die, nach einem mehrjährigen obligatorischen Physikunterricht in der Sekundarstufe I, auf ein anhaltendes fachliches Interesse hinweist. Beide Probendenauswahlkriterien werden als präventive Maßnahme zur Begrenzung der Probandenheterogenität verstanden.

Die Ziehung der Stichprobe wird auf das regionale und überregionale Umfeld des Schülerlaborstandortes beschränkt, weil eine landesweite Stichprobenwahl aus ökonomischen und organisatorischen Gründen wenig erfolgversprechend erscheint. Ausgewählt werden Physikkurse aus Bochum, Witten, Mülheim (Ruhr), Lüdenscheid und Essen, deren Kursteilnehmer und -lehrer vor Beginn der Studie die Bereitschaft zum Besuch des Schülerlabors bekundet haben. Die Bereitschaft zum Schülerlaborbesuch ist eine notwendige Voraussetzung für das Stichprobenauswahlverfahren, um einen Stichprobenumfang mit möglichst geringer Ausfallquote zu gewährleisten. Das Auswahlverfahren basiert auf einer gezielten, nicht-zufälligen Auswahl von Physikkursen, deren Gruppenzugehörigkeit (Test- und Kontrollgruppe) durch Randomisierung bestimmt wird. Um schulspezifischen Effekten zu begegnen, werden für jede Test- bzw. Kontrollgruppe zwei Physikkurse ermittelt. Für die Bearbeitung der offenen Experimentieraufgabe werden die Probanden aufgefordert, durch Selbstregulation Probandenpaare zu bilden. Dies hat zur Konsequenz, dass mit dem Messinstrument Beobachtungsbogen die Experimentierprozesse nicht von einzelnen Probanden, sondern von Probandenpaaren abgebildet werden. Im Schülerlabor werden von 77 Probandenpaaren (154 Probanden) die Beobachtungsdaten erhoben. Die Gesamtzahl aller Teilnehmer an der Studie beträgt 181. Der Ausfall liegt im Durchschnitt bei zwei Schülern pro Physikkurs. Zwei Kurse zeichnen sich durch den vierfachen Ausfall aus, weil den Kursteilnehmern am Datenerhebungstag (Beobachtung der Experimentierprozesse) ein Besuch im Schülerlabor aus schulinternen Gründen (Klausurtermin) nicht stattgegeben wurde.

Die auf alle Teilnehmer bezogene Ausfallquote erhöht sich dadurch auf 14 %. Die „Rücklaufquote“ ist trotzdem sehr zufriedenstellend.

Teilstichprobe. Zur Auswertung der Beobachtungsdaten wird aus der Stichprobe eine Teilstichprobe gezogen. Sie umfasst 41 Probandenpaare (82 Probanden). Die Probanden haben ein durchschnittliches Alter von 16 Jahren und 2 Monaten. Die Teilstichprobe setzt sich aus 52 Schülern und 30 Schülerinnen zusammen. Das naturwissenschaftliche Profil, welches aus dem Kurswahlverhalten der Probanden abgeleitet wird, zeichnet sich durch einen hohen naturwissenschaftlichen Anteil aus: Zwei Drittel der Probanden der Teilstichprobe haben sich in der Jahrgangsstufe 11 neben Physik für eine weitere Naturwissenschaft (vor allem Biologie) und ein Sechstel der Probanden hat sich neben Physik für zwei weitere Naturwissenschaften entschieden. Ein Sechstel der Probanden der Teilstichprobe hat für die gymnasiale Oberstufe Physik als einzige Naturwissenschaft bestimmt.

Die Ziehung der Teilstichprobe erfolgt nach zwei Kriterien, die die Probandenpaare durch Selbststeuerungsprozesse nach Abschluss der offenen Experimentieraufgabe erfüllt haben müssen. Die Kriterien garantieren die Parallelisierung der Teilstichprobe (vgl. Kap. 4.3). Entgegen einer Parallelisierung durch externe Regulierung, die häufig zum Zeitpunkt der Studienplanung eingeleitet wird, unterliegt die Parallelisierung durch Selbststeuerung dem Zufallsprinzip (zufallsverknüpfte Parallelisierung). Nach Rost bieten Zufallsstichproben (hier: *Zufallsteilstichprobe*) den Vorteil, dass sie eine Verallgemeinerung der Ergebnisse auf die Probandenpopulation erlauben (Rost, 2007, S. 88).

Das erste Auswahlkriterium der Teilstichprobe betrifft das Experimentierverhalten. Alle Probandenpaare der Teilstichprobe stimmen darin überein, dass sie beim Experimentieren die naturwissenschaftliche Arbeitsweise Umgang mit Variablen nicht nur im obligatorischen Experimentierteil (Profil konstruieren, Experimentierphase 2), sondern auch im optionalen Experimentierteil (Profilkonstruktion vorbereiten, Experimentierphase 1) gezeigt haben müssen. Zur optionalen Experimentierphase der offenen Experimentieraufgabe zählen beispielsweise die Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelsets und/ oder die Planung von eigenen Profilkonstruktionen. Als mittlere Experimentierzeit pro Probandenpaar konnte für beide Experimentierphasen zusammen eine Zeit von 140 Minuten bestimmt werden. Gruppenspezifische Unterschiede lagen diesbezüglich nicht vor. Das zweite Auswahlkriterium betrifft die Dimension naturwissenschaftliche Arbeitsweise (NAW) des schriftlichen Messinstruments Fach- & NAW-Test zum Messzeitpunkt 1 (Vortest). Für die Auswahl der Teilstichprobe werden nur Probandenpaare zugelassen, deren Teilnehmer zum Messzeitpunkt 1 in der Dimension naturwissenschaftliche Arbeitsweisen gleiche bzw. ähnliche Testergebnisse erzielt haben. Das hat zur Konsequenz, dass sich die einzelnen Probanden der Probandenpaare hinsichtlich

ihrer allgemeinen Fähigkeit zum Umgang mit Variablen – gemessen an dem Gültigkeitsbereich des schriftlichen Testinstruments – nicht unterscheiden.

Die Ergebnisse des Fach- & NAW-Tests können für die Dimension naturwissenschaftliche Arbeitsweisen (NAW) zum Messzeitpunkt 1 im Kapitel 4.3 und im Anhang eingesehen werden (Abbildung 7, Tabelle 7, Abbildung A 4). Über die Gruppenvergleiche hinaus werden die probandenpaarspezifischen Ergebnisse im Kapitel 4.5 sowie im Anhang differenziert wiedergegeben (Abbildungen A 5 a – c). An der Zusammensetzung der Test- und Kontrollgruppen kann zudem erkannt werden (Abbildung A 4 bzw. A 5a), dass eine gleichmäßige Verteilung der Probandenpaaranzahlen nach den Ergebnissen des Vortests (Messzeitpunkt 1, Dimension NAW) über den oberen, mittleren und unteren Punktebereich beabsichtigt wurde. Die gewählte Verteilung über die drei Punktebereiche bietet den Vorteil, dass die Teilstichprobe nicht von Extremgruppen besetzt wird, von denen bekannt ist, dass ihre Testergebnisse bei wiederholenden Messungen zu Mittelwerten regredieren (Rost, 2007, S. 100 ff). Für den oberen, mittleren und unteren Punktebereich werden je vier Probandenpaare gezogen, was pro Test- bzw. Kontrollgruppe zu einer Probandenpaaranzahl von 12 (bzw. 24 Probanden) führt. Alle Auswahlkriterien können für die Testgruppen Reproduktion (TG 1), die Testgruppe Transfer (TG 3) und die Kontrollgruppe (KG) erfüllt werden.

Die Testgruppe Reorganisation (TG 2) unterscheidet sich von den anderen Experimentalgruppen zum einen durch die Gesamtzahl der Probandenpaare (5) und zum anderen durch die Herkunft der Probanden. Während die Probandenpaare der Testgruppe 1 (Reproduktion), der Testgruppe 3 (Transfer) und der Kontrollgruppe aus zwei Physikkursen verschiedener Gymnasien stammen, besuchen die Probandenpaare der Testgruppe 2 (Reorganisation) denselben Physikkurs. Das hat zur Folge, dass die Ergebnisse von einem nicht auflösbaren schulspezifischen Effekt begleitet werden können. Ferner hat die gegenüber den anderen Testgruppen verminderte Probandenpaaranzahl zur Folge, dass die Probandenpaare der Testgruppe 2 (Reorganisation) über den oberen, mittleren und unteren Punktebereich inhomogen verteilt sind, was z. B. an den Konfidenzintervallen der Abbildung 7 oder an der Verteilung der Ergebnisse des NAW-Tests zum Messzeitpunkt 1 der Abbildung A 4 abgelesen werden kann. Gegenüber der Testgruppe 1 (Reproduktion), der Testgruppe 3 (Transfer) und der Kontrollgruppe haben die Unregelmäßigkeiten in der Gruppenzusammensetzung zur Konsequenz, dass die Ergebnisse der Testgruppe 2 (Reorganisation) nur mit eingeschränkter Gültigkeit interpretiert werden können.

Zur Ziehung der Teilstichprobe sei abschließend angemerkt, dass nach Erfüllung aller Kriterien zwischen einigen Probandenpaaren Auswahlmöglichkeiten bestanden. In diesen Fällen wurden die Messdaten von Probandenpaaren ausgewertet,

deren Experimentiertätigkeiten von Datenaufnehmern erhoben wurden, die sich durch eine höhere Beobachterübereinstimmung ausweisen konnten (vgl. Kap. 4.2.3).

Variablen. *Unabhängige Variablen (UV).* Die Gruppenzugehörigkeit der Probanden ist die unabhängige Variable. Unterschieden wird zwischen der Kontrollgruppe (KG) und den drei Testgruppen (TG 1 – 3) zwischen denen der Kontext der vorbereitenden Instruktion variiert wird.

Abhängige Variablen (AV). Der Umgang mit Variablen beim Experimentieren mit bzw. ohne vorbereitende Instruktion zählt zu den abhängigen Variablen. Das Konstrukt Umgang mit Variablen wird durch das Variablenkonzept festgelegt bzw. definiert (vgl. Kap. 3). Die Erhebung des Konstrukts erfolgt mit dem Messinstrument Beobachtungsbogen. Eine weitere abhängige Variable wird unter dem Konstrukt allgemeine Fähigkeiten im Umgang mit Variablen zu den Messzeitpunkten 1 – 3 auf Basis eines schriftlichen Fach- & NAW-Tests erfasst (vgl. Kap. 4.2.2).

Kontroll- und Störvariablen. Die abhängigen Variablen werden unter Kontrolle des Fachwissens erhoben. Mit der Strategie der Konstanthaltung, der Strategie der Parallelisierung und der Strategie der Randomisierung werden potentielle Störvariablen wie der Einfluss der Lehrervariablen auf die vorbereitende Instruktion, die zeitliche Dauer der vorbereitenden Instruktion, die Experimentierbedingungen, die Laborvariablen des Schülerlabors, die Schulform, die Kurszugehörigkeit und das Kurswahlverhalten der Probanden sowie das Alter und das regionale Einzugsgebiet (Ruhrgebiet) der Probanden neutralisiert. Eine potentielle Störvariable, die nicht neutralisiert wurde, ist der zeitliche Abstand zwischen der vorbereitenden Instruktion und der Bearbeitung der offenen Experimentieraufgabe, der in Abhängigkeit der Test- und Kontrollgruppen zwischen drei und sieben Tagen variiert. Dieser potentielle Störfaktor ist dem Anspruch geschuldet, dass zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe kein Physikunterricht stattfinden soll, der vom Kurslehrer erteilt wird. Der Einfluss der Störvariablen „variierender zeitlicher Abstand“ wird als gering eingeschätzt. Demgegenüber wird der Physikunterricht, den der Fachlehrer in den drei Monaten zwischen den Messzeitpunkten 2 und 3 (Nachtest und Verzögerter Nachtest) erteilt, als eine einflussreichere Störvariable bewertet. Der Störfaktor kann im Rahmen der Studie nicht ausgeschaltet werden.

4.2.2 Messinstrument Fach- & NAW-Test

Dieses Kapitel wird durch fünf Abschnitte gegliedert. Im ersten Abschnitt wird das Konstrukt „Allgemeine Fähigkeiten im Umgang mit Variablen“ vorgestellt, dass mit dem Testinstrument erhoben werden soll. Der zweite Abschnitt bezieht sich auf den

Aufbau und die Struktur des Messinstruments, der dritte Abschnitt auf die Bewertung der Testaufgaben, der vierte Abschnitt auf die Gütekriterien des Messinstruments und im fünften Abschnitt werden kritische Anmerkungen zum Fach- & NAW-Test gemacht. Mit einem Vorgriff auf die Datenauswertung (Kap. 4.3, Kap. 4.5) kann schon jetzt berichtet werden, dass mit dem Einsatz des Messinstruments mehrere Ziele verfolgt werden. Hierzu gehören die Parallelisierung der Experimentalgruppen (KG, TG 1 – 3), die Kontrolle des Fachwissens sowie ein Vergleich der Testergebnisse des NAW-Tests mit den direkt beobachteten Ergebnissen des Experimentierens (Niveaustufen des Experimentierens). Weiterhin wird das Messinstrument benutzt, um die Entwicklung der allgemeinen Fähigkeiten für den Umgang mit Variablen zu dokumentieren.

Der Namensgebung folgend, handelt es sich bei dem Fach- & NAW-Test um ein zweidimensionales Messinstrument, mit dem das physikalische Fachwissen (1. Dimension) und die naturwissenschaftlichen Arbeitsweisen (2. Dimension) erfasst werden. Das physikalische Wissen wird mit dem Fach-Test geprüft. Die meisten Testitems können dem Sachgebiet Mechanik zugeordnet werden mit denen das allgemeine physikalische Wissen im Sachgebiet und das spezielle physikalische Wissen in ausgewählten Themengebieten erhoben wird. Zu den ausgewählten Themengebieten zählen: statischer Auftrieb, dynamischer Auftrieb, Energie und Energieumwandlung. Mit dem NAW-Test werden die naturwissenschaftlichen Arbeitsweisen erfasst. Der Schwerpunkt der Dimension liegt in der Messung des Konstrukts „Allgemeine Fähigkeiten im Umgang mit Variablen“. Da das Konstrukt durch Literaturquellen nicht spezifiziert werden kann, folgt zunächst eine Konstruktbeschreibung, die weder als umfassend noch als abgeschlossen eingestuft werden darf; sie hat einen einführenden Charakter. Hierbei ist wichtig zu wissen, dass die Konstruktbeschreibung nicht empirisch gewonnen, sondern nach einer Analyse des Themenfeldes normativ erstellt wurde.

Beschreibung des Konstrukts „Allgemeine Fähigkeiten im Umgang mit Variablen“. Das Konstrukt „Allgemeine Fähigkeiten im Umgang mit Variablen“ wird in zwei Fähigkeitsbereiche unterteilt, die durch mehrere Teilfähigkeiten definiert werden. Zum ersten Fähigkeitsbereich mit den Teilfähigkeiten 1 – 4 zählen elementare Fähigkeiten im Umgang mit Variablen. Zum zweiten Fähigkeitsbereich mit den Teilfähigkeiten 5 – 7 zählen weiterführende Fähigkeiten im Umgang mit Variablen. In der Tabelle 2 werden die beiden Fähigkeitsbereiche mit ihren korrespondierenden Teilfähigkeiten dargestellt.

Elementarer Fähigkeitsbereich. Der Tabelle 2 ist zu entnehmen, dass die elementaren Fähigkeiten das Erkennen von relevanten Variablen (Teilfähigkeit 1), das Testen und Kontrollieren von relevanten Variablen (Teilfähigkeit 2) und den bewertenden Vergleich von getesteten Variablen (Teilfähigkeit 3) beinhalten.

Zu den Teilfähigkeiten des ersten Fähigkeitsbereichs werden folgende Anmerkungen gemacht:

Die erste Teilfähigkeit (relevante Variablen erkennen) umfasst die Identifikation von potentiellen Testvariablen aus einer vorgegebenen Faktorenkombination und die vorgabenfreie Identifikation von potentiellen Testvariablen (eigenständige Suche potentieller Testvariablen).

Zur zweiten Teilfähigkeit (relevante Variablen testen und kontrollieren) werden die explizite Kenntnis von Experimentierstrategien, die Fähigkeit zur Anwendung von Experimentierstrategien und die Fähigkeit zum Wechsel von Experimentierstrategien gezählt. Als Beispiele zweier Experimentierstrategien seien die Trial- & Error-Strategie und die Variablenkontrollstrategie genannt. Die Fähigkeit zum Strategiewechsel wird als wichtig erachtet, weil sie für den Fall, dass eine gewählte Experimentierstrategie bei der Aufgabenlösung nicht zum Erfolg führt, eine Handlungsalternative für den Lernenden bietet.

Zu der dritten Teilfähigkeit (getestete Variablen vergleichen und bewerten) gehört, dass für verschiedene Variablenverläufe funktionale Zusammenhänge beschrieben, mögliche Intervallbegrenzungen erkannt und Aussagen über Effektstärken von Variablen gemacht werden können. Als Beispiele seien Variablen mit linearem, extremalem und konstantem (ohne Einfluss) Verlauf genannt. Eine notwendige Voraussetzung der Teilfähigkeit ist, dass die Qualität der Messdaten vom Datenauswerter beurteilt werden kann. Dazu müssen nicht nur Kenntnisse über den Datenaufnahmeprozess, sondern auch Grundkenntnisse im Umgang mit Messunsicherheiten vorhanden sein. Die Forderungen illustrieren, dass die Auswertung von Messwerten nicht vom Datenaufnahmeprozess entkoppelt werden kann.

Über die Teilfähigkeiten 1 – 3 hinaus ist eine weitere Teilfähigkeit für Datenauswerter von Bedeutung, die die auszuwertenden Messdaten nicht von Experimentatoren übergeben bekommen, sondern in experimentellen Lernumgebungen eigenständig generieren. Der Anspruch, experimentelle Daten aufnehmen und auswerten zu können, erfordert prozessuale Fähigkeiten, die als vierte Teilfähigkeit im elementaren Fähigkeitsbereich berücksichtigt werden. Zu den prozessualen Fähigkeiten im Umgang mit Variablen gehört, dass der Zusammenhang zwischen Variablen erkennen (Teilfähigkeit 1), Variablen testen und kontrollieren (Teilfähigkeit 2) und Variablen vergleichen und bewerten (Teilfähigkeit 3) hergestellt und als wiederholt zu durchlaufende Abfolge verstanden wird.

Weiterführender Fähigkeitsbereich. Der weiterführende Fähigkeitsbereich des Konstrukts beinhaltet (vgl. Tabelle 2), dass der Umgang mit Variablen in verschiedenen Kontexten und Domänen beherrscht wird (Teilfähigkeit 5), dass der Umgang mit Variablen für Aufgabenstellungen mit unterschiedlichen Komplexitätsgraden angewendet werden kann (Teilfähigkeit 6) und dass wissenschaftliche

Methoden der Erkenntnisgewinnung genutzt werden können, die für den Umgang mit Variablen relevant sind (Teilfähigkeit 7).

Zu der Teilfähigkeit 5 des weiterführenden Fähigkeitsbereichs ist anzumerken: Die Auswahl der Kontexte soll nicht auf naturwissenschaftliche Domänen wie Biologie, Chemie und Physik beschränkt werden, sondern auch naturwissenschaftsverwandte Domänen wie Ingenieurwissenschaften, Medizin und Pharmazie mit einbeziehen. Die Fähigkeit, auch ohne Kontextbindung mit Variablen umgehen zu können (Dekontextualisierung), wird als eine Transferleistung der allgemeinen Fähigkeiten für den Umgang mit Variablen eingestuft. Eine Dekontextualisierung wird nur für die Datenauswertung (Messfolgen von Variablen vergleichen und bewerten) und nicht für die Datenaufnahme (Variablen erkennen, testen und kontrollieren) als sinnstiftend erachtet, da in den naturwissenschaftlichen und naturwissenschaftsverwandten Domänen (fast ausschließlich) Messdaten aufgenommen werden, die an Kontexte gebunden sind.

Zu der Teilfähigkeit 6 des weiterführenden Fähigkeitsbereichs ist anzumerken, dass die Komplexität von Aufgabenstellungen hinsichtlich des Umgangs mit Variablen durch drei Stufen graduert wird. Zu der untersten Komplexitätsstufe gehört, dass mit vorgegebenen und bereitgestellten Experimentiermaterial Messfolgen aufgenommen und ausgewertet werden können (Beispiel: Mit dem bereitgestellten Experimentiermaterial „Flügelset“ Messfolgen testen und auswerten). Zu der mittleren Komplexitätsstufe gehört, dass aus vorgegebenem und bereitgestelltem Experimentiermaterial Messfolgen neu zusammengestellt, aufgenommen und ausgewertet werden können (Beispiel: Aus dem bereitgestellten Experimentiermaterial „Flügelset“ neue Messfolgen durch Kombination von Einzelprofilen zusammenstellen und die Messfolgen testen und auswerten). Zu der obersten Komplexitätsstufe gehört, dass aus nicht vorgegebenem und bereitgestelltem Experimentiermaterial Messfolgen neu erstellt, aufgenommen und ausgewertet werden können (Beispiel: Profile für das Modell eines Windrades eigenständig erstellen, die Profile als Messfolgen zusammenstellen und die Messfolgen testen und auswerten). Die drei Komplexitätsstufen unterscheiden sich durch die Anzahl der möglichen Lösungen (eine oder mehrere) und die Anzahl der möglichen Lösungswege (einer oder mehrere). Es sei darauf verwiesen, dass die Anzahl der möglichen Lösungen und Lösungswege in dieser Arbeit auch als Prüfkriterium für den Öffnungsgrad von Aufgaben verstanden wird, weshalb mit einem höheren Komplexitätsgrad auch ein höherer Öffnungsgrad einhergeht. Wie der Beschreibung der Komplexitätsstufen entnommen werden kann, sind alle Komplexitätsgrade mit den Teilfähigkeiten des elementaren Fähigkeitsbereichs „Variablen erkennen“, „Variablen testen und kontrollieren“ sowie „Variablen vergleichen und bewerten“ verbunden.

Zu der Teilfähigkeit 7 des weiterführenden Fähigkeitsbereichs sind folgende Anmerkungen zu treffen: Die Auswahl von wissenschaftlichen Methoden der Erkenntnisgewinnung wird dadurch beschränkt, dass die Methoden für den Umgang mit Variablen relevant sein sollen. Zu den relevanten wissenschaftlichen Methoden der Erkenntnisgewinnung werden die induktive Methode, die deduktive Methode sowie die Verschränkung (der gegenseitige Zusammenhang) von induktiver und deduktiver Methode gezählt. Zu bedenken ist, dass die genannten Methoden mit weiteren naturwissenschaftlichen Methoden und Arbeitsweisen wie explorieren, beobachten, Hypothesen aufstellen, Experimente planen, idealisieren, Hypothesen überprüfen usw. verbunden sein können.

Fähigkeitsbereiche (FB) und Teilfähigkeiten des Konstrukts „Allgemeine Fähigkeiten im Umgang mit Variablen“	
FB	Beschreibung der Teilfähigkeit
1	Variablen erkennen (1.)
1	Variablen testen und kontrollieren (2.)
1	Variablen vergleichen und bewerten (3.)
1	prozessuale Fähigkeiten, Zusammenhang zwischen Teilfähigkeiten 1 – 3 herstellen (4.)
2	Umgang mit Variablen in verschiedenen Kontexten und Domänen (5.)
2	Umgang mit Variablen für Aufgaben mit unterschiedlichem Komplexitätsgrad (6.)
2	Anwendung wissenschaftlicher Methoden der Erkenntnis- gewinnung, die für den Umgang mit Variablen relevant sind (7.)
Legende: 1 Elementarer Fähigkeitsbereich 2 Weiterführender Fähigkeitsbereich 1. – 7. Nummer der Teilfähigkeit	

Tab. 2: Fähigkeitsbereiche und Teilfähigkeiten des Konstrukts
„Allgemeine Fähigkeiten im Umgang mit Variablen“

Wie die vorausgehenden Beschreibungen verdeutlichen, handelt es sich bei den allgemeinen Fähigkeiten im Umgang mit Variablen um ein komplexes Konstrukt.

Für den Aufbau und die Struktur eines schriftlichen Messinstruments hat das zur Konsequenz, dass unter Berücksichtigung von Testgütekriterien alle sieben Teilfähigkeiten erhoben werden müssen. Infolgedessen wäre ein valides Messinstrument zu entwickeln, mit dem die aufgestellten Fähigkeitsbereiche überprüft werden können. Da die Testkonstruktion und Testanalyse nicht zu den Zielen der vorliegenden Studie gehören, werden stattdessen Items aus gut dokumentierten und mehrfach erprobten Messinstrumenten entnommen und als Testaufgaben für den NAW-Test verwendet. Das Verfahren hat eine Minderung der Testgüte zur Konsequenz, die gemessen an den Zielen der Untersuchung – begründete Hypothesen zu generieren – akzeptiert wird. Weiterhin sei angemerkt, dass aus denselben Gründen auf eine Pilotierung des Messinstruments verzichtet wird.

Aufbau und Struktur des Messinstruments. *NAW-Test.* Die PISA-Studie 2006, die TIMS-Studie der Population 2 und die TIMS-Studie der Population 3 dienen als Itemquelle des NAW-Tests. Die Auswahl der Testaufgaben aus den publizierten Itempools erfolgt nach mehreren Kriterien; sie legen den Aufbau und die Struktur des Messinstruments fest. Ein Auswahlkriterium der Testaufgaben für den NAW-Test ist die Kategorisierung und Klassifizierung der Items durch die Autoren. Beispielsweise werden die Aufgaben im Naturwissenschaftstest 2006 der PISA-Studie nach Kompetenzen klassifiziert, die zu drei Aufgabentypen mit verschiedenen Varianten führen (Hammann, 2006a, 2006b; Prenzel et al., 2007). Nach der Testkonzeption sind alle drei Aufgabentypen dem Kompetenzbereich „Prozesse und Prozeduren [...], die von den Schülerinnen und Schülern verstanden und beherrscht werden sollten“ zugeordnet (Prenzel et al., 2007, S. 38). Eine Analyse der Aufgabentypen ergibt, dass Teilfähigkeiten, die mit dem NAW-Test erhoben werden sollen nur mit Items des ersten und dritten Aufgabentyps und nicht mit Items des zweiten Aufgabentyps gemessen werden können. Zum ersten Aufgabentyp der PISA-Studie 2006 gehört das „Erkennen naturwissenschaftlicher Fragestellungen und Verstehen naturwissenschaftlicher Untersuchungen“ zum zweiten Aufgabentyp das „Erklären naturwissenschaftlicher Phänomene“ und zum dritten Aufgabentyp das „Umgehen mit naturwissenschaftlicher Evidenz.“ Ein weiteres Kriterium für die Auswahl der Testaufgaben ist die Itemschwierigkeit. Zur Charakterisierung der Testaufgaben aus der PISA-Studie werden die Kompetenzstufen benutzt, die den Aufgabenbeispielen zugeordnet werden. Zur Charakterisierung der Testaufgaben aus den TIMS-Studien werden die internationalen Schwierigkeitsindizes verwendet, die für die Aufgabenbeispiele erstellt wurden. Für die Zusammenstellung des NAW-Tests wird die selbst gewählte Vorgabe erfüllt, Items aus beiden Quellen mit verschiedenen Schwierigkeitsindizes und Kompetenzstufen zu kombinieren. Das dritte Kriterium, welches insbesondere bei der Itemauswahl aus den TIMS-Studien angewendet wird, ist, dass Aufgaben mit konstruktionsbedingten Defiziten möglichst nicht verwendet werden. Denn als ein Fazit fachlicher und fachdidak-

tischer Analysen konnte festgestellt werden, dass in den TIMS-Studien Testaufgaben mit Mängeln hinsichtlich der fachlichen Richtigkeit, hinsichtlich der sprachlichen Verständlichkeit und hinsichtlich der Eindeutigkeit der Fragestellung eingesetzt wurden (Duit, 1999; Kühnelt, 1999; Hagemeister, 1999). Das vierte und letzte Kriterium betrifft die Testlänge, die nach der Definition von Lienert & Raatz (1998, S. 33) der Anzahl der Testaufgaben entspricht. Sie hängt u. a. vom Grad der Homogenität bzw. Heterogenität des zu erfassenden Merkmalbereichs und vom Grad der erwünschten Reliabilität des Tests ab. Ungeachtet dessen wird die Testlänge des Fach- & NAW-Tests durch die Praktikabilität und durch die Motivationslage der Probanden festgelegt, welche nach Mossbrugger & Kelava (2008, S. 34 f.) nicht vernachlässigt werden darf. Denn die Genauigkeit der Messung kann durch den Einfluss von testfremden Variablen, wie ein Absinken der Konzentration und Motivation, beeinträchtigt werden. Aus diesem Grund wird die Testlänge durch die Testzeit begrenzt, die eine Schulstunde mit einer Dauer von 45 Minuten nicht überschreiten soll. Die festgelegte Bearbeitungszeit für die insgesamt 24 Testitems des Fach- & NAW-Tests (11 Items Fach-Test, 13 Items NAW-Test) ist identisch mit der Testdauer, die den Probanden zur Lösung des TIMS-Naturwissenschaftstests (Population 2) zur Verfügung gestellt wurde. Aus dieser zeitlichen Vorgabe wurde in Verbindung mit der mittleren Bearbeitungszeit eines TIMS-Items die Anzahl aller Testitems für den Fach- & NAW-Test bestimmt. Nach den Angaben der Literatur lag die mittlere Bearbeitungszeit pro TIMS-Item bei 80 Sekunden (Kühnelt, 2000). Mit demselben Zeitmaß wurden auch die Items, die nicht aus den TIMS-Studien stammen bemessen (s. u.).

Die Herkunft der Testaufgaben verteilt sich wie folgt: Von den 13 Items des NAW-Tests wurden sechs Items aus zwei verschiedenen Aufgabeneinheiten der Pisa-Studie entnommen. Hierbei handelt es sich um die Aufgabeneinheiten Sonnenschutz und Windenergie der PISA-Beispielaufgaben 2006. Aus der TIMS-Studie der Population 2 wurden drei Items entnommen (Baumert et al., 1998, Items I 12, N 1, P 7). Damit möglichst viele Teilfähigkeiten der allgemeinen Fähigkeiten im Umgang mit Variablen erfasst werden können, wurde die genannte Itemauswahl durch eine selbst entwickelte Aufgabeneinheit mit der thematischen Ausrichtung statischer Auftrieb (Item SA), die aus vier Testitems mit halboffenem Antwortformat besteht, ergänzt. Dennoch gelingt es nach einer normativen Einschätzung nicht, dass mit dem NAW-Test alle Teilfähigkeiten des Konstrukts erhoben werden. Insbesondere wird vermutet, dass die Teilfähigkeit 6 (Umgang mit Variablen für Aufgaben mit unterschiedlichem Komplexitätsgrad) nur im Ansatz und die Teilfähigkeit 7 (Anwendung wissenschaftlicher Methoden der Erkenntnisgewinnung) gar nicht erfasst wird. Zudem wird eingeschätzt, dass die höchste Komplexitätsstufe der Teilfähigkeit 6 (Messfolgen aus nicht vorgegebenem und bereitgestelltem Experimentiermaterial neu erstellen, testen und auswerten) mit keinem Item erreicht wird.

Weiterhin wird auf eine differenzierte Zuweisung der Testaufgaben zu den Teilfähigkeiten verzichtet, weil mit den TIMS-Studien, der PISA-Studie und der vorliegenden Studie zum Teil konträre Ziele verfolgt werden. Stattdessen werden die Items einem Sub-Fähigkeitsbereich, der aus mehreren Teilfähigkeiten des elementaren und weiterführenden Fähigkeitsbereichs besteht, zusammengefasst. Der Schwerpunkt, der mit den ausgewählten TIMSS- und PISA-Items gemessen wird, liegt in einem (Sub-) Fähigkeitsbereich, der von der Teilfähigkeit 2 (Variablen testen und kontrollieren), von der Teilfähigkeit 3 (Variablen vergleichen und bewerten) und von der Teilfähigkeit 5 (Umgang mit Variablen in verschiedenen Kontexten und Domänen) dominiert wird.

Obwohl nach einer normativen Einschätzung mit dem NAW-Test nicht alle Teilfähigkeiten des Konstrukts abgebildet werden, wird er als ein geeignetes Messinstrument eingestuft, dass dem explorativ-quantitativen Charakter der hypothesengenerierenden Studie gerecht wird. Für eine hypothesenüberprüfende Untersuchung würde eine umfassendere Entwicklung des Testinstruments notwendig.

Fach-Test. Während die Items des Fachtests aus denselben Quellen wie die Items des NAW-Tests stammen, werden die Itemauswahlkriterien des NAW-Tests auf den Fach-Test übertragen und darüber hinaus um ein weiteres Kriterium ergänzt: Die Itemauswahl des Fach-Tests wurde zudem auf Basis der Richtlinien und Lehrpläne der gymnasialen Oberstufe Physik (Jgst. 11, G 9, NRW) mit den fachinhaltlichen Lernvoraussetzungen der Probanden abgestimmt. Das Kriterium wurde besonders bei der Auswahl der Test-Items aus der TIMS-Studie (Population 3) beachtet. Die Herkunft der Testaufgaben verteilt sich wie folgt: Von den 11 Items des Fach-Tests wurden zwei Items aus einer Aufgabeneinheit der Pisa-Studie entnommen. Hierbei handelt es sich um die Aufgabeneinheit Windenergie der PISA-Beispielaufgaben 2006. Aus der TIMS-Studie der Population 2 wurden drei Items entnommen (Items K 17, L 1, O 13). Aus der TIMS-Studie der Population 3 wurden fünf Items entnommen (Baumert et al., 1999, Items B 6, H4, G 11, G 15, Y 1), wovon ein Item als Anker-Item in beiden Populationen verwendet wurde (Item Y 1). Die genannte Itemauswahl wird durch eine selbst entwickelte Aufgabe mit einem offenen Antwortformat zum Thema dynamischer Auftrieb ergänzt (Item DA).

Bewertung der Testaufgaben des Fach- & NAW-Tests. Zur Bewertung der Itemantworten werden die Codierungsschlüssel der TIMS- und PISA-Studien angewendet (vgl. TIMSS Population 2 Item Pool, TIMSS Population 3 Item Pool, PISA-Lösungen der Beispielaufgaben 2006). Die Lösungen der Items bzw. Aufgabeneinheiten, die selbst entwickelt wurden, werden auf Grundlage von Musterlösungen und Bewertungsschlüsseln begutachtet. Die Auswertung der Fach- & NAW-Tests erfolgt numerisch. Die Punktierung (Vergabe der Punkte) orientiert sich an der Punktierung der TIMSS-Items (Baumert et al., 1998) sowie an der mittleren

Bearbeitungszeit pro TIMSS-Item (Population 2). Sie wird mit einem Punkt als Basisgröße der Bewertung für einen Zeittakt von 80 Sekunden Testzeit festgelegt, was zu einer Gesamtpunktzahl von 26 Punkten für die Bearbeitung von 24 Items und zu einer berechneten Gesamttestdauer von 35 Minuten führt (13 Punkte Fach-Test, 13 Punkte NAW-Test). Die verbleibende Testzeit wurde zum Lesen des Stimulus-Materials der Aufgabeneinheiten Sonnenschutz, statischer Auftrieb und Wind-energie veranschlagt. Die Bewertung bzw. Punktierung der Testaufgaben ist – zusammen mit der Struktur des Messinstruments, dem Aufbau des Messinstruments und den selbst entwickelten Aufgabenstellungen – im Anhang dokumentiert.

Angemerkt sei, dass die mittlere Bearbeitungszeit bzw. Punktierung pro Testitem, die als Basisgröße der Testkonstruktion zur Berechnung der Gesamttestdauer diene, nicht als starres, sondern als dynamisches Zeitmaß verwendet wurde. Die errechnete Bearbeitungszeit wurde nicht zu kurz bemessen, weil die tatsächliche von den Probanden der Stichprobe beanspruchte mittlere Gesamttestdauer für den ersten Messzeitpunkt 36 Minuten betrug. In dieser Zeit ist das Lesen des Stimulus-Materials bereits enthalten.

Wie in der Einleitung des Kapitels berichtet wurde, wird das Messinstrument benutzt, um die Probanden der Experimentalgruppen (KG, TG 1 – 3) auf Basis ihrer Ergebnisse im NAW-Test zu parallelisieren. Zu diesem Zweck wird für den NAW-Test ein oberer, mittlerer und unterer Punktebereich geschaffen (vgl. Kap. 4.3 ff.). Die Einweisung in die Punktebereiche erfolgt nach einer prozentualen Vorschrift. Die Grenzen der prozentualen Zuordnungsvorschrift müssen dynamisch verwendet werden, weil die Ziehung der Teilstichprobe in Kombination mit einem zweiten Auswahlkriterium, dem Experimentierverhalten der Probandenpaare, durchgeführt wird (Kap. 4.2.1).

Gütekriterien des Messinstruments. Obwohl für die TIMS- und PISA-Studien Testgütekriterien bestimmt worden sind (z. B. $0,80 \leq \alpha \leq 0,90$, Baumert et al., 1998, S. 11; z. B. $0,76 \leq \alpha \leq 0,79$, Baumert et al., 1999, S. 15), ist eine Übernahme der testtheoretischen Kennwerte für den Fach- & NAW-Test, der mehrheitlich aus Items der zitierten Quellen besteht, nicht zulässig. Auf eine erneute Bestimmung der Testgüte wird dennoch verzichtet, weil konstruktionsbedingte Mängel der TIMSS-Items nicht durch hohe Reliabilitäten (α) kompensiert werden können. In Analogie dazu darf an der PISA-Testkonzeption bezweifelt werden, dass mit den eingesetzten schriftlichen Messinstrumenten prozessuale naturwissenschaftliche Kompetenzen gemessen werden können. Ferner werden für die selbst entwickelten Items weder Schwierigkeitsindizes noch Kompetenzstufen ermittelt. Deshalb wurde die Anzahl der selbst konstruierten Testaufgaben und Aufgabeneinheiten minimiert. Aus demselben Grund wurde entschieden, die Teilfähigkeit 6 nur im Ansatz und die Teilfähigkeit 7 gar nicht zu messen. Die Entscheidung hat zur Konsequenz, dass der

Grad der Heterogenität des erfassten Konstrukts- bzw. Merkmalbereichs und in dessen Folge die Testlänge des Fach- & NAW-Tests beschränkt wird.

Kritische Anmerkungen zum Aufbau und zu den Gütekriterien des Messinstruments. Der Fach- und NAW-Test ist ein schriftliches Messinstrument, dessen Items aus mehreren Testinstrumenten bezogen und durch selbst konstruierte Aufgaben ergänzt wurden. Allerdings dürfen die testtheoretischen Kennwerte der Bezugsquellen wie z. B. das Gütekriterium der Normierung für den Fach- & NAW-Test nicht übernommen werden. Alternativ hätten mehrfach erprobte und in der Literatur gut dokumentierte Tests mit akzeptablen Testgütekriterien wie TIPS (Test of Integrated Science Processes, Burns et al., 1985) eingesetzt werden können, ohne dass die angesetzte Testdauer von 45 Minuten hätte verändert werden müssen. Obwohl die Gütekriterien des TIPS-Tests überzeugend sind, wurde er als NAW-Test nicht verwendet. Denn zum einen hat die Analyse eines Itemauszugs dazu geführt, dass mehrere Kritikpunkte der schriftlichen Testpraxis (Stebler et al., 1998, S. 29) für diesen Test zutreffend sind. Zum anderen darf bezweifelt werden, dass mit dem Messinstrument das Konstrukt „Allgemeine Fähigkeiten im Umgang mit Variablen“ valider gemessen werden kann als durch das praktizierte Verfahren der gezielten Itemauswahl und Itemkomposition. Eine Validierung des Konstrukts wurde für diese hypothesengenerierende Studie, aus der begründete aber keine allgemein gültigen Aussagen gewonnen werden, nicht durchgeführt. Es sei darauf verwiesen, dass für einige Ziele, die mit dem Einsatz des Fach- & NAW-Tests verfolgt werden, wie z. B. die Parallelisierung der Experimentalgruppen (KG, TG 1 – 3), eine Validierung des Konstrukts nicht notwendig ist und den explorativ-quantitativen Anspruch der Untersuchung übererfüllen würde.

4.2.3 Messinstrument Beobachtungsbogen

Das Kapitel *Messinstrument Beobachtungsbogen* wird durch drei Abschnitte gegliedert. Der erste Abschnitt bezieht sich auf die Struktur, den Aufbau und den Gebrauch des Messinstruments, der zweite Abschnitt auf die Codierung der mit dem Messinstrument erhobenen Experimentierhandlungen und der letzte Abschnitt auf die Gütekriterien der Beobachtung und der Codierung. Die Gütekriterien sind sowohl für die Beobachtung als auch für die Codierung zu bestimmen, weil der kriteriengeleitete Datenaufnahmeprozess durch die Person des Beobachters vom kriteriengeleiteten Datenauswertungsprozess durch die Person des Codierers entkoppelt wurde.

Struktur, Aufbau und Gebrauch des Messinstruments. Die Strukturierung und der Aufbau des Beobachtungsinstruments basieren auf den Erkenntnissen, die in der Vorstudie (vgl. Kap. 3) über den Beobachtungsprozess generiert wurden. Das

Ergebnis der Entwicklungsarbeit ist ein kategoriengeleitetes, halbstandardisiertes Messinstrument, mit dem für den Umgang mit Variablen prozess- und produktbezogene Daten des Experimentierens erfasst werden (Kirchner & Priemer, 2010). Um die Experimentierabläufe zu beschreiben, werden mit dem Messinstrument von geschulten Beobachtern die beobachtbaren Experimentiertätigkeiten (5 Hauptkategorien) und verbalen Argumentationen (3 Hauptkategorien) der Probandenpaare erhoben. Dem Beobachter steht eine Methodenvielfalt (Methodentriangulation) zur Verfügung: Er registriert die sichtbaren, beobachtbaren Experimentiertätigkeiten und die verbalen Selbstauskünfte (Argumentationen) der Probanden, die aus dem kommunikativen Austausch der Probandenpaare (Partnerarbeit) hervorgehen. Sie finden in Begleitung der beobachtbaren Experimentiertätigkeiten statt. Zudem hat der Beobachter die Möglichkeit, die Probanden im begrenzten Umfang zu befragen.

Struktur des Messinstruments. Die Strukturierung des Beobachtungsinstruments wird durch drei Merkmale erreicht. Das erste Strukturmerkmal betrifft eine Unterscheidung zwischen den beiden Experimentierphasen Profilkonstruktion vorbereiten (Experimentierphase 1) und Profil konstruieren (Experimentierphase 2). Die grundlegende Ausrichtung des Messinstruments nach den Experimentierphasen ergibt sich aus den Anforderungen, die an die Bearbeitung der offenen Experimentieraufgabe gestellt werden. Das zweite Merkmal der Strukturierung betrifft die Verteilung der Hauptkategorien der beobachtbaren Experimentiertätigkeiten auf die beiden Phasen. Von den fünf Hauptkategorien gehören die ersten drei zur Phase Profilkonstruktion vorbereiten (Experimentierphase 1) und die letzten beiden zur Phase Profil konstruieren (Experimentierphase 2). Die fünf Hauptkategorien sind: 1. Information beschaffen, 2. Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset (inkl. Testung), 3. Planung einer Profilkonstruktion, 4. Profil konstruieren (inkl. Testung) und 5. Änderung an einer bestehenden Profilkonstruktion (inkl. Testung).

Zu den verbalen Argumentationen gehören folgende drei Hauptkategorien: 1. Vergleich und Bewertung von Messdaten bzw. Messfolgen (V), 2. Strategie und Idee zur Profilkonstruktion (S) und 3. Erklärung (E, z. B. fachliche Erklärung). Anders als an den Kategorien der beobachtbaren Experimentiertätigkeiten, kann an den Kategorien der verbalen Argumentation keine experimentierphasenspezifische Zuordnung erkannt werden, denn sie sind für die beiden Experimentierphasen identisch. Eine Unterscheidung zwischen den verbalen Argumentationen in Phase 1 und den verbalen Argumentationen in Phase 2 kann dennoch getroffen werden, weil die Kategorien der verbalen Argumentationen in Begleitung von beobachtbaren Experimentiertätigkeiten auftreten. Die Zuordnung der verbalen Argumentationen zu den beobachtbaren Experimentiertätigkeiten erfolgt über eine Zeitmarke.

Das dritte Merkmal betrifft die zeitliche Strukturierung des Beobachtungsbogens. Sie ermöglicht eine zeitintervallbasierte Auflösung der Beobachtungen. Dazu werden die Beobachtungsereignisse, die den Hauptkategorien zugeordnet werden können, nach einem Zeitstichprobenverfahren (Bortz & Döring, 2002, S. 272) in 5-Minuten Intervalle sequenziert. Die Wahl des Zeitintervalls beruht auf den Erkenntnissen, die in den explorativen Vorstudien (Kap. 3) generiert wurden und ist den beobachtbaren Experimentiertätigkeiten angepasst. Eine zeitliche Auflösung der verbalen Argumentationen wird nicht verfolgt. Die Abbildung 5 zeigt die Strukturierung des Messinstruments durch die Merkmale Experimentierphase, Hauptkategorie und Experimentierzeit.

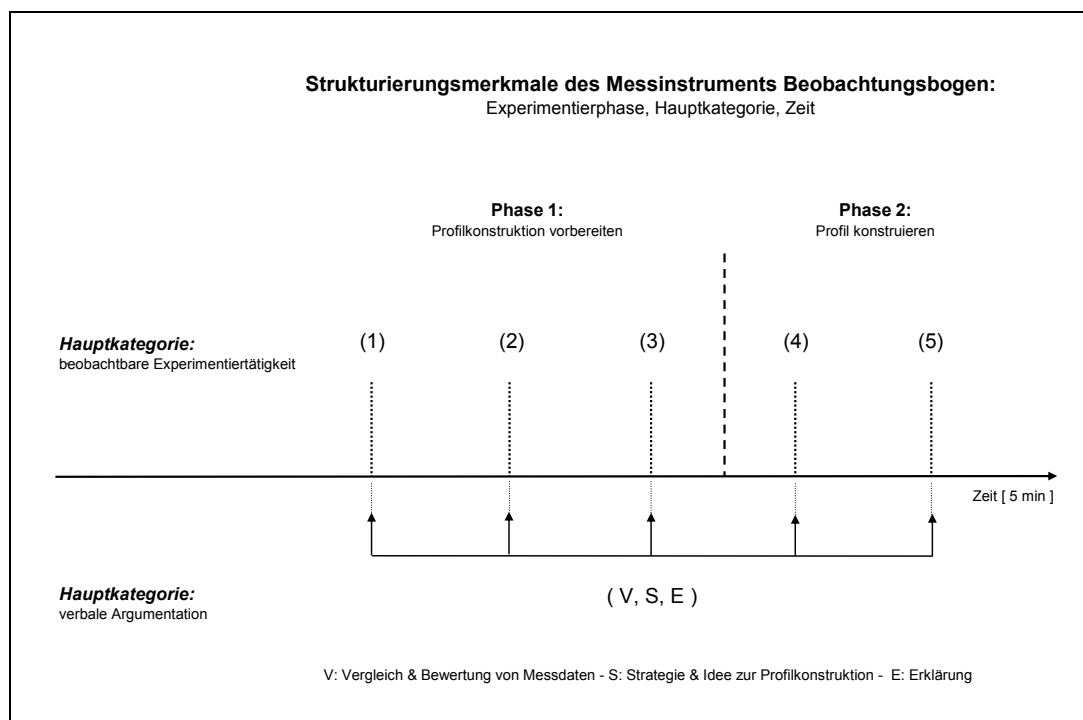


Abb. 5: Strukturierung des Messinstruments Beobachtungsbogen

Aufbau des Messinstruments. Beim Aufbau bzw. bei der Gestaltung des Beobachtungsinstruments muss berücksichtigt werden, dass für den Protokollant auch dann eine kategoriengeleitete Mitschrift der Beobachtungen möglich ist, wenn Probanden sprunghaft und in schneller Folge einen Wechsel der Experimentiertätigkeiten ausführen. In Abhängigkeit der Experimentierphase und der Kategorie konnten im Rahmen der explorativen Vorstudien (Kap. 3) Experimentiersequenzen identifiziert werden, in denen die Informationsdichten erhöht sind. Die von den Probanden gelieferten und von den Beobachtern zu protokollierenden Informationen sind vor allem für die beobachtbare Experimentiertätigkeit der Kategorie Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset erhöht (Experimentierphase 1, Profilkonstruktion vorbereiten). Gegenüber der Phase 2 sind die

Unterschiede in der Informationsdichte damit zu erklären, dass für die Arbeit mit den Flügelsets die Profile bereitgestellt werden, wohingegen in der Experimentierphase 2, die Profile von den Probanden selbst konstruiert werden müssen. Wenn Probanden ihre Aktivitäten auf der Ebene der beobachtbaren Experimentiertätigkeiten erhöhen, dann hat das auch Auswirkungen auf die verbalen Argumentationen der Kategorie „Vergleich und Bewertungen von Messfolgen“. Sie kommen in der Phase 1 gehäuft vor als in der Phase 2, vor allem bei Probandenpaaren der Testgruppen (TG 1 – 3), die vor der Bearbeitung der offenen Experimentieraufgabe eine vorbereitende Instruktion erhalten haben. Um in Experimentiersequenzen mit hohen Informationsdichten die Gefahr eines möglichen Datenverlustes während der Aufnahme des Beobachtungsprozesses zu minimieren, wurde das Messinstrument hinsichtlich des Aufbaus und hinsichtlich der Protokollführung ökonomisiert. Für die Phase 2 hat sich eine kategoriengeleitete, freie Protokollführung und für die Phase 1 eine kategoriengeleitete, halbstandardisierte Protokollführung als geeignet bewährt. Am Beispiel ausgewählter Experimentiersequenzen wird der prinzipielle Aufbau des Beobachtungsinstruments vorgestellt und diskutiert, zunächst für die erste und anschließend für die zweite Experimentierphase.

Aufbau des Messinstruments. Experimentierphase 1. Ein sprunghafter und schneller Wechsel von Experimentiertätigkeiten liegt beispielsweise vor, wenn Probanden bei der Arbeit mit den Flügelsets mit der Aufnahme einer nicht konfundierenden Messfolge beginnen, ihre Experimentiertätigkeit konfundierend durch Kombination mehrerer Profile aus verschiedenen Sets fortsetzen und abschließend Vergleiche zwischen Messwerten vornehmen, die nicht konfundierenden und konfundierenden Messfolgen entnommen wurden. Für den Protokollanten ist die schriftliche Aufnahme der Beobachtungen ein Problem, da er während des laufenden Experimentierprozesses die Entwicklung von Experimentiertätigkeiten nicht vorhersehen kann. Dadurch steigt insbesondere bei wechselhaften Experimentiertätigkeiten die Gefahr von Fehlzuordnungen. Gelöst wird die Problematik durch die Gestaltung des Beobachtungsbogens. Im Messinstrument ist für die Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset deshalb eine Sammlung von Ereignissen und Elementarereignissen zu finden, wohingegen Subkategorien, wie die konfundierende/ nicht konfundierende Aufnahme von Messfolgen dem Wortlaut nach im Beobachtungsbogen explizit nicht mit aufgenommen worden sind. Die Protokollführung wird so festgelegt, dass der Protokollant seine Beobachtungen den bereits aufgelisteten Ereignissen bzw. Elementarereignissen zuordnet und in Abhängigkeit der zeitlichen Anordnung durch eine Markierung (Strich) im Beobachtungsbogen bekundet. Über die Einordnung der Ereignisse und Elementarereignisse in Subkategorien (z. B. nicht konfundierende/ konfundierende

Messfolge) hat ein Codierer erst nach Abschluss des Beobachtungsverfahrens zu entscheiden.

Durch die Trennung des Beobachtungsprozesses vom Codierprozess (Eingruppierung der Ereignisse und Elementarereignisse in Subkategorien) gewinnt der Beobachter an Flexibilität, so dass er auch in Phasen mit hoher Informationsdichte auf wechselhafte Experimentiertätigkeiten reagieren und sich auf die Selektion der Beobachtungsobjekte und auf das Abfassen des Protokolls konzentrieren kann. Wieder am Beispiel der Arbeit mit den Flügelsets, soll die Gestaltung des Beobachtungsbogens durch Ereignisse und Elementarereignisse illustriert werden (vgl. Anhang, Beobachtungsbogen, S. 3 f.). Die Arbeit mit den Flügelsets setzt sich aus einem Dreischritt der Ereignisse Transport (eines Flügelsets zum Arbeitsplatz), Montage (eines Profils aus dem gewählten Flügelset am Generator) und Test (des gewählten Profils durch Anströmen) zusammen, der durch weitere Elementarereignisse weiter spezifiziert wird. Für die Profiltestung gehören dazu die Elementarereignisse Flügel anströmen, Messdaten aufnehmen, Experimentierbedingungen/ Kontrollbedingungen einhalten, experimenteller Misserfolg und experimentelle Fehlersuche (vgl. Anhang, Beobachtungsbogen, S. 3). Um davon die Arbeit mit Flügelsetkombinationen (Information für den Codierer: Es kann sich um konfundierende Experimente handeln) zu unterscheiden, wird in dem Dreischritt aus Transport, Montage und Test nur das Elementarereignis „Montage eines Profils“ aus dem gewählten Flügelset gegen die „Montage einer Profilkombination“, die von den Probanden aus Profilen verschiedener Flügelsets zusammengesetzt werden kann, ausgetauscht (vgl. Anhang, Beobachtungsbogen, S. 4). Die im Beobachtungsbogen auf derselben Seite weiter unten aufgeführten Elementarereignisse zur Testung bleiben von Veränderungen unberührt, beziehen sich nun aber auf die Arbeit mit Flügelsetkombinationen. Der Protokollant hat dadurch die Flexibilität, auch auf sprunghafte Wechsel im Experimentierverhalten reagieren zu können. Darüber hinaus wurde der Beobachtungsbogen so gestaltet, dass für jedes Profil eines Flügelsets und für jede Flügelsetkombination der Eintritt eines Ereignisses (Transport, Montage, Test) festgestellt und schriftlich schnell fixiert werden kann. Dieses Verfahren der Protokollführung wurde für die Experimentierphase 1 optimiert und zwar für alle Hauptkategorien der beobachtbaren Experimentiertätigkeiten und für die Kategorie der verbalen Argumentation „Vergleich und Bewertung von Messungen bzw. Messfolgen“. Die weiteren Kategorien der verbalen Argumentation (Strategie und Idee zur Profilbildung, Erklärungen) treten im Vergleich zur Kategorie „Vergleich und Bewertung“ weniger gehäuft auf, so dass sich eine freie Protokollierung unter der Rubrik Anmerkung als die ökonomischere Protokollführung erwiesen hat. Die Rubrik Anmerkung steht den Protokollanten auch dann zur Verfügung, wenn beobachtete Experimentier-

tätigkeiten den aufgeführten Ereignissen und Elementarereignissen nicht zugewiesen werden können.

Aufbau des Messinstruments. Experimentierphase 2. Im Gegensatz dazu hat sich für die zweite Experimentierphase für alle Kategorien der beobachtbaren Experimentiertätigkeiten und den verbalen Argumentationen eine kategoriengeleitete, freie Protokollführung als die geeignetere Variante bewährt. Das Resultat basiert auf den Ergebnissen der explorativen Vorstudien (vgl. Kap. 3) die zeigten, dass zwischen den Probandenpaaren erhebliche Unterschiede im Experimentierablauf auftraten, wenn eigene Profile konstruiert werden mussten. Die Diskrepanzen zwischen den Probandenpaaren waren besonders groß, wenn sie vor der Bearbeitung der offenen Experimentieraufgabe keine Vorbereitung (KG) bzw. eine vorbereitende Instruktion (TG 1 – 3) erhalten haben. Eine der Phase 1 analoge Gestaltung des Messinstruments in Phase 2 wäre deshalb nur unter Verlust von Beobachtungsdaten realisierbar gewesen. Trotz der freien Protokollführung wird der Beobachter in der Phase 2 durch eine Auswahl von Ereignissen und Elementarereignissen auf die zu protokollierenden Experimentierhandlungen fixiert, so dass der Systematisierungsgrad der Beobachtung von dem Wechsel der Protokollführung unberührt bleibt. Die Selektion der Beobachtungsereignisse ist wie in Phase 1 kategoriengeleitet. Dies erfolgt mit Hilfe eines zweiseitigen Beiblattes (vgl. Anhang, Beobachtungsbogen, S. 8 f.), das durch sieben Spalten strukturiert ist. Der Beobachter hat die Aufgabe mittels des Beiblatts, eine kategoriengeleitete Beobachtungsmitschrift zu erstellen und zwar unter Einhaltung des zeitlichen Eintritts der Ereignisse. Die Intervalllänge des Zeitstichprobenverfahrens beträgt wieder fünf Minuten.

Im Unterschied zur Experimentierphase 1 sind auf dem Beiblatt die Ereignisse und Elementarereignisse aufgenommen, die zu der vierten und fünften Hauptkategorie der beobachtbaren Experimentiertätigkeiten gehören (Profil konstruieren, Änderung an einer bestehenden Profilkonstruktion). Wiederholt mitaufgelistet sind die Ereignisse zum Testen von Profilen und die Ereignisse, die zu den Kategorien der verbalen Argumentationen gehören. Sie unterscheiden sich inhaltlich nicht von denen der Phase 1. Durch die phasenübergreifende Verknüpfung von Kategorien wird auf dem Beobachtungsinstrument die Gesamtzahl verschiedener Ereignisse und Elementarereignisse deutlich reduziert.

Gebrauch des Messinstruments. Der Gebrauch bzw. Einsatz des Messinstruments erfolgt in einer nicht-teilnehmenden, nicht verdeckten Beobachtermethode. Pro Probandenpaar wird das experimentelle Vorgehen von einem Beobachter protokolliert. Wie andere offene Datenerhebungsmethoden, zählt die angewandte Beobachtungsmethode nicht zu den nonreaktiven Verfahren. Um die Beeinflussbarkeit des Beobachtungsprozesses durch den Kontakt des Datenaufnehmers mit den Datengebern zu kontrollieren, wird der Beobachtungsvorgang von Beobachtungsvorschriften begleitet. Dazu gehören auch Regeln, die bei einer Befragung der

Probanden einzuhalten sind. Der Beobachter darf beispielsweise nicht als Informationsquelle für die Probanden fungieren und mögliche Nachfragen des Beobachters dürfen sich nur auf bereits ausgeführte Tätigkeiten der Probanden beziehen. Das Ziel der Maßnahmen ist, den Grad der Beobachterreaktivität zu senken. Fragen der Probanden können jederzeit an einen Projektbetreuer gestellt werden. Sie werden vom Protokollanten in der Mitschrift notiert und markiert.

Das Messinstrument wird von 12 Beobachtern benutzt. Ein hoher Grad an Systematisierung der Beobachtung wird nicht nur durch die Konstruktion eines strukturierten, sondern auch durch die kontrollierte Anwendung dieses Messinstruments bedingt. Die Forderung nach einer adäquaten Abbildung des Forschungsobjektes (Experimentierprozesse) ist erfüllt, wenn die Datenaufnehmer zu einer differenzierten Wahrnehmung und zu einem differenzierten Bericht hinsichtlich des Forschungsziels und des Forschungsplans sensibilisiert bzw. befähigt worden sind. Mit dem Ziel, den Beobachtern eine systematische Beobachtung zu ermöglichen, so dass die Selektion der Beobachtungsobjekte nicht dem Ermessensspielraum des Beobachters überlassen bleibt, ist ein Beobachtertraining über die Struktur, den Aufbau und den Gebrauch des Messinstruments notwendig. Minimiert wird dadurch die Gefahr, dass der Beobachter den Beobachtungsgegenstand selektiv verzerrt, was zu Beobachterfehlern führen würde. Gänzlich vermieden werden könnte dieser Effekt z. B. durch eine videografische Aufzeichnung der Beobachtungsprozesse. Die Methode stellt aber keine Alternative dar, weil die meisten Probanden den Einsatz von Videokameras nicht akzeptierten. Die Probanden entzogen sich dem Sichtfeld der Kamera und schränkten die verbalen Argumentationen ein, so dass eine Aufnahme von Experimentierprozessen mit erheblichen Beobachtungseinbußen verbunden gewesen wäre. Stattdessen wurde unter Einbindung verschiedener Beobachtungstechniken (Friedrichs, 1980, S. 274) für die Datenaufnehmer ein Beobachtertraining entwickelt, in dem verschiedene Schulungsmaßnahmen vereinigt worden sind (Pinther, 1980; Bortz & Döring, 2002, S. 273). Das Training ist zeitlich auf sechs Stunden ausgelegt und fand an drei Tagen in zweistündigen Schulungen statt. Um Fehlerquellen durch Beobachtereffekte wie Hypothesenbefangenheit o. ä. zu vermeiden, gehört eine Aufklärung der Beobachter über die Forschungsfragen nicht zu den Inhalten der Schulung.

Aufbau, Struktur und Anwendung des Codiersystems. Zur Aufbereitung der erhobenen Experimentierhandlungen wird ein Codiersystem erstellt, mit dessen Hilfe die Ereignisse und Elementarereignisse des Experimentierens in Prozessdiagramme überführt werden können. Das entwickelte Codiersystem basiert auf zwei Codierprinzipien und ist auf die Operationalisierung der Lernziele zum Umgang mit Variablen ausgerichtet. Das betrifft vor allem die Codierung von Messfolgen, woran die Tragfähigkeit des Codiersystems überprüft werden kann. Zur Abbildung der beobachtbaren Experimentiertätigkeiten wird ein numerisches und

zur Abbildung der verbalen Argumentationen („Tonspur“) ein symbolisches Codierprinzip verwendet. Der Einsatz verschiedener Verschlüsselungsprinzipien bietet den Vorteil, dass mit der Codierung zugleich die Methode der Datenerhebung transportiert wird. Beide Codierprinzipien sind aufeinander abgestimmt, so dass durch die Kombination von numerischen und symbolischen Codes detaillierte Experimentierhandlungen verschlüsselt werden können. Bei der Beschreibung der Kategorien und Subkategorien wurde bedacht, dass eine Operationalisierung der Lernziele für den Umgang mit Variablen nicht nur auf das gewählte Aufgabenbeispiel „Konstruktion von Windradflügel“ angewendet werden kann.

Codierung beobachtbarer Experimentiertätigkeiten. Numerisches Codierprinzip. In Analogie zum Aufbau des Beobachtungsbogens, ist das numerische System zur Codierung der beobachtbaren Experimentiertätigkeiten auf eine Unterscheidung zwischen den Experimentierphasen Profilkonstruktion vorbereiten (Experimentierphase 1) und Profil konstruieren (Experimentierphase 2) ausgerichtet. Die Phasenzugehörigkeit ist an der ersten Ziffer einer Codierung zu erkennen, die die Tätigkeiten der Probanden auf der Durchführungsebene des Experimentierens anzeigt. Sie bestimmt die Hauptkategorie. Mit Ausnahme der Kategorie „Profilkonstruktion planen“, deren Inhalte nur durch eine Codierung mit Hilfe verbaler Argumentationen weiter spezifiziert wird, charakterisiert eine Folge von Subkategorien die beobachtbaren Experimentiertätigkeiten. Jede Subkategorie wird durch eine neue Ziffer codiert, so dass eine Experimentiersequenz durch eine Ziffernfolge numerisch verschlüsselt wird. Die Anzahl der Ziffern bestimmt den Grad der Spezifizierung einer Experimentiersequenz.

Mit einer Ziffernfolge von bis zu vier Ziffern (Hauptkategorie und drei Subkategorien) ist der Grad der Systematisierung besonders hoch, wenn Probanden Messdaten bzw. Messfolgen erheben. Um eine spätere Decodierung der Prozessgrafiken zu erleichtern, ist die Positionierung einer Ziffer in der Ziffernfolge an eine Systematik gebunden. Die ersten beiden Ziffern verschlüsseln die Tätigkeiten der Probanden. Die erste Ziffer legt die Hauptkategorie fest. Die zweite Ziffer steht für eine Spezifizierung der durch die Hauptkategorie angezeigten Tätigkeit der Probanden. Die dritte Ziffer enthält in Abhängigkeit von der Hauptkategorie entweder Informationen über die Qualität einer Experimentierstrategie (nicht konfundierend/ konfundierend) oder über die Qualität einer Arbeitsweise (messdatenbasiert/ nicht messdatenbasiert). Die letzte Ziffer einer vierstelligen Ziffernfolge spezifiziert entweder eine Tätigkeit der Probanden zur Bereitstellung von Messdaten, wie die schriftliche Datenaufnahme oder die Präzision, mit der eine Tätigkeit umgesetzt wird. Dazu zählt z. B. die Kontrolle der Ausprägung einer Variablen.

1. Information beschaffen	
Code	Bedeutung der Codierung
1.1	Informationsquelle
1.2	... Infomappen
1.3	... Internet
1.4	... weiteres Probandenpaar/ Nachbargruppe
1.5	... schriftliches Material der Vorbereitung
	... Flügelsets
2. Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset	
Code	Bedeutung der Codierung
2.1.1.1	Messfolge aufnehmen, nicht konfundierend,
2.1.1.2	... Messdaten schriftlich notieren
2.1.2.1	... Messdaten nicht schriftlich notieren
2.1.2.2	Messfolge aufnehmen, konfundierend,
2.2.1.1	... Messdaten schriftlich notieren
2.2.1.2	... Messdaten nicht schriftlich notieren
2.2.2.1	Einzelmessung aufnehmen, konfundierend,
2.2.2.2	... Messdaten schriftlich notieren
	... Messdaten nicht schriftlich notieren
3. Planung einer Profilkonstruktion	
Code	Bedeutung der Codierung
3	Planung einer/ mehrerer Profilkonstruktionen

Tab. 3a: Numerisches Codiersystem zur Codierung beobachtbarer Experimentiertätigkeiten in Experimentierphase 1

4. Profil konstruieren	
Code	Bedeutung der Codierung
4.1.1	Profilkonstruktion durch Parameterkombination ... messdatenbasiert
4.1.2	... nicht messdatenbasiert
4.2.1	Profilkonstruktion durch Kopie eines Flügelsetprofils ... messdatenbasiert
4.2.2	... nicht messdatenbasiert
Profilkonstruktion testen	
Code	Bedeutung der Codierung
	Profilkonstruktion testen, Messdaten schriftlich notieren
	Profilkonstruktion testen, Messdaten nicht schriftlich notieren
	Profilkonstruktion nicht testen, Messdaten nicht schriftlich notieren
5. Änderung an einer bestehenden Profilkonstruktion	
Code	Bedeutung der Codierung
5.1.1.1	eine Änderung vornehmen, nicht konfundierend, ... mit Kontrolle der Ausprägung
5.1.1.2	... ohne Kontrolle der Ausprägung
5.1.2.1	eine Änderung vornehmen, konfundierend, ... mit Kontrolle der Ausprägung
5.1.2.2	... ohne Kontrolle der Ausprägung
5.2.1.1	wiederholt Änderung vornehmen, nicht konfundierend, ... mit Kontrolle der Ausprägung
5.2.1.2	... ohne Kontrolle der Ausprägung
5.2.2.1	wiederholt Änderung vornehmen, konfundierend, ... mit Kontrolle der Ausprägung
5.2.2.2	... ohne Kontrolle der Ausprägung

Tab. 3b: Numerisches Codiersystem zur Codierung beobachtbarer Experimentiertätigkeiten in Experimentierphase 2

Ebenfalls mit dem Ziel, eine Decodierung zu erleichtern, folgt die Gestaltung des Codiersystems einer weiteren Systematik. Tätigkeiten, Experimentierstrategien oder Arbeitsweisen werden mit der Zahl 1 und die komplementären Tätigkeiten, Experimentierstrategien oder Arbeitsweisen werden mit der Zahl 2 codiert (Aufnahme einer Messfolge/ Einzelmessung, nicht konfundierend/ konfundierend, messdatenbasierte/ nicht messdatenbasierte Profilkonstruktion). Am Beispiel der Aufnahme einer Messfolge mit dem bereitgestellten Experimentiermaterial Flügelsets (Experimentierphase 1, Fallbeispiel 1) und durch Änderungen an einer bestehenden Profilkonstruktion (Experimentierphase 2, Fallbeispiel 2) soll exemplarisch der Aufbau des Codiersystems erklärt werden.

Fallbeispiel 1: Zur Codierung wird eine Experimentiersequenz ausgewählt, nach der die Probanden mit dem bereitgestellten Experimentiermaterial Flügelset eine Messfolge mit allen Profilen des Sets unter Anwendung einer nicht konfundierenden Experimentierstrategie aufnehmen und die Messdaten schriftlich protokollieren. Die Experimentiersequenz wird durch die vierstellige Ziffernfolge 2.1.1.1 codiert. Die Codierung ist der Tabelle 3a entnommen.

Fallbeispiel 2: Für die Codierung einer Messfolge, die von den Probanden durch eigene Profilkonstruktionen (Fallbeispiel 2a) oder Änderungen an einer bestehenden Profilkonstruktion aufgenommen werden (Fallbeispiel 2b), ist auf Basis des numerischen Codiersystems eine Aneinanderreihung von mehreren Ziffernfolgen notwendig. Zur Codierung vorgestellt wird eine Experimentiersequenz, nach der die Probanden eine messdatenbasierte Profilkonstruktion durch Kombination mehrerer Parameter erstellen, nach Fertigstellung des Profils eine Variable unter Anwendung einer nicht konfundierenden Experimentierstrategie verändern und bei der Einstellung der Variablen die Ausprägung der Variablen kontrollieren. Anschließend wird von den Probanden unter Anwendung derselben Experimentierstrategie, wieder unter Kontrolle der Ausprägung, dieselbe Variable zweimal verändert (Codierung mit einer vierstelligen Ziffernfolge je Wiederholung). Diese Experimentiersequenz wird durch eine Verkettung der vier Ziffernfolgen (4.1.1), (5.1.1.1), (5.2.1.1), (5.2.1.1) codiert. Jede einzelne der vierstelligen Ziffernfolge ist der Tabelle 3b entnommen.

Angemerkt sei erstens, dass die Tätigkeit des schriftlichen Protokollierens von Messdaten bei der Codierung einer Messfolge, die durch eigene Profilkonstruktionen erstellt wurde, in der Beschreibung der Kategorie nicht mit aufgenommen wurde (vgl. vierte und fünfte Hauptkategorie). Das war notwendig, weil Probanden bei eigenen Profilkonstruktionen wiederholt den Produktions- vom Messprozess zeitlich entkoppelt haben. Zur Erstellung von Prozessdiagrammen wurde deshalb für die Testung von Profilkonstruktionen eine eigenständige, symbolische Codierung gewählt, die eine flexible Codierung der Tätigkeit erlauben (vgl. Tabelle 3b). Zweitens, dass aus der Kombination der beiden letzten Ziffern abzulesen ist, welche

Akzente die Probanden bei der Aufnahme von Messfolgen setzten. Die Kombination eröffnet die Möglichkeit zur Beantwortung von Fragen wie „Kontrollieren die Probanden bei der Einstellung von Variablen an einer bestehenden Profilkonstruktion die Ausprägung der Variablen, obwohl die Experimentierstrategie konfundierend ist?“ (Codierung: 5.1.2.1), oder zeigen die Probanden ein Experimentierverhalten, durch das die umgekehrte Fragestellung beantwortet werden kann („Wenden die Probanden bei Aufnahme einer Messfolge durch Änderung einer Profilkonstruktion eine nicht konfundierende Experimentierstrategie an, ohne dass die Ausprägung der Variablen bei ihrer Einstellung kontrolliert wird?“; Codierung 5.1.1.2). Festgestellt werden kann auch, wenn Probandenpaare ein wechselndes Experimentierverhalten zeigen, so dass keine der Fragestellungen eindeutig zu beantworten wäre.

Die Tabellen 3a und 3b fassen unter Angabe der Haupt- und Subkategorien die numerischen Codierungen beobachtbarer Experimentiertätigkeiten für die erste und zweite Experimentierphase zusammen.

Codierung verbaler Argumentationen. Symbolisches Codierprinzip. Anders als das numerische Codiersystem, ist das symbolische Codiersystem zur Verschlüsselung verbaler Argumentationen nicht mehr auf eine der beiden Experimentierphasen zugeschnitten. Notwendig sind von den Experimentierphasen unabhängige Codiersymbole, weil jede Kategorie der beobachtbaren Experimentiertätigkeit potentiell von jeder Kategorie der verbalen Argumentation begleitet werden kann. Die Tabellen 4a – c zeigen Übersichten von allen drei Hauptkategorien der verbalen Argumentationen, die durch mehrere Subkategorien weiter aufgeschlüsselt werden. Ihre Codes markieren im Prozessdiagramm verbale Argumentationen der Kategorie „Vergleich und Bewertung von Messfolgen (V)“, der Kategorie „Strategie und Idee zur Profilkonstruktion (S)“ und der Kategorie „Erklärungen (E)“. Die Kategorien unterliegen nur zum Teil einer hierarchischen Ordnung.

Die vier Subkategorien der Kategorie „Vergleich und Bewertung von Messfolgen (V)“ unterliegen von V_1 nach V_4 einer aufsteigenden, streng hierarchischen Ordnung (Tabelle 4a).

Die Subkategorien der Kategorie „Strategie und Idee zur Profilkonstruktion (S)“ (Tabelle 4b) unterliegen von S_1 nach S_4 nur zum Teil einer aufsteigenden hierarchischen Ordnung, denn für die Bewertung der Experimentierprozesse ist die hinter der Strategie stehende messdatenbasierte/ nicht messdatenbasierte Arbeitsweise relevant. Die Strategie S_1 kommt nur in der nicht messdatenbasierten und die Strategie S_4 nur in der messdatenbasierten Variante vor, so dass zwischen ihnen eine hierarchische Ordnung besteht. Empirisch konnten für die Strategien S_2 und S_3 beide Ausprägungen der Arbeitsweise bestimmt werden.

Code	Kategorien der verbalen Argumentation Vergleich und Bewertung von Messfolgen (V)
V ₁	Schlussfolgerung auf eine Flügeleigenschaft durch Bewertung eines Messergebnisses einer Einzelmessung ohne Vergleich mit anderen Messergebnissen („Bewertung ohne Vergleich“) und/ oder Vergleich konfundierender Messergebnisse/ Messfolgen mit Bewertung und Schlussfolgerung auf die Flügeleigenschaft
V ₂	Vergleich und Bewertung eines Messergebnisses mit einem anderen und Schlussfolgerung auf die bessere Flügeleigenschaft („paarweise Vergleich mit Bewertung“)
V ₃	Vergleich und Bewertung von Messergebnissen einer Messfolge und Schlussfolgerung auf die Flügeleigenschaft, z. B. Verlauf einer Variablen, Effektstärke einer Variablen
V ₄	Vergleich und Bewertung von Messergebnissen mehrerer Messfolgen und Schlussfolgerung auf die Flügeleigenschaft, z. B. Verlauf und Effektstärken von Variablen, Rangfolge von Maximalwerten

Tab. 4a: Kategorien der verbalen Argumentation „Vergleich und Bewertung von Messfolgen“

Code	Kategorien der verbalen Argumentation Strategie und Idee zur Profilkonstruktion (S)
S ₁	Willkürliche ad hoc Profilkonstruktion mit oder ohne Parameternennung
S ₂	Profilkonstruktion durch Kopie eines Profils, z. B. aus der Messfolge eines Flügelsets oder einer erstellten Profilkonstruktion
S ₃	Profilkonstruktion durch Vereinigung von Bedingungen zur Erzeugung von maximalen Spannungswerten mehrerer Variablen
S ₄	Profilkonstruktion durch Vereinigung von Bedingungen zur Erzeugung von maximalen Spannungswerten mehrerer Variablen (vgl. S ₃) unter Berücksichtigung der Effektstärke von Variablen und Rangfolgenbildung der Maximalwerte

Tab. 4b: Kategorien der verbalen Argumentation „Strategie und Idee zur Profilkonstruktion“

Code	Kategorien der verbalen Argumentation Erklärung (E)
E ₁	Plausibilitätserklärung, z. B. ästhetische Begründung für Wahl einer Profilkonstruktion, Begründung durch Wissen aus dem Alltag
E ₂	Vermutung zu potentiellen Messergebnissen vor Durchführung eines Experiments, z. B. zur Effektstärke einer Variablen, zum Verlauf einer Variablen, zur Ausprägung einer Variablen
E ₃	Fachliche Erklärung oder Begründung z. B. zur Abhängigkeit/ Unabhängigkeit einer Variablen (hier: von der Drehzahl), zur Vernachlässigung eines Messergebnisses

Tab. 4c: Kategorien der verbalen Argumentation „Erklärung“

Von den Subkategorien der Kategorie „Erklärung (E)“ folgen nur die Subkategorien E₁ und E₃ einem hierarchisch aufsteigenden Prinzip (Tabelle 4c). In den explorativen Vorstudien (Kap. 3) konnte festgestellt werden, dass Vermutungen zu potentiellen Messergebnissen, die vor der Durchführung eines Experiments geäußert wurden (E₂), nach der Durchführung des Experiments von den Probanden nicht mehr aufgegriffen werden. Das heißt, die Subkategorie kann als singuläre, verbale Argumentation auftreten. Des Weiteren konnten Kopplungen der Subkategorie E₂ mit den Subkategorien E₁ und E₃ gefunden werden, so dass die vor einem Experiment geäußerten Vermutungen nach Abschluss des Experiments durch Plausibilitätserklärungen (E₁) oder durch fachliche Erklärungen (E₃) wieder aufgegriffen werden.

In Anlehnung an die hierarchische Ordnung der Subkategorien E₁ und E₃, folgen die Kopplungen der Argumentationen E₂/ E₁ und E₂/ E₃ demselben hierarchischen aufsteigenden Prinzip.

Kombination des numerischen und symbolischen Codierprinzips. Für die Codierung von Experimentiersequenzen werden Synergieeffekte genutzt, die sich aus der Kombination beider Codierprinzipien ergeben, mit dem Ziel, die Anzahl der Ziffern einer numerischen Codierung auf maximal vier Ziffern zu begrenzen. Dazu werden zwei Beispiele vorgestellt, die nur durch Anwendung des aufgestellten numerischen oder des symbolischen Codierprinzips nicht zu erfassen wären.

Beispielsweise kann bei der Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset durch die numerische Codierung einer Messfolge nicht unterschieden werden, ob für die Messfolge nur zwei Messwerte (Ansatz einer Messfolge) oder mehr als zwei Messwerte aufgenommen wurden. Die Kombination

aus der numerischen Codierung einer Messfolge (z. B. Codierung 2.1.1.1) mit der symbolischen Codierung V_2 ("paarweiser Vergleich") zeigt an, dass mit einem Flügelset nur eine Messfolge im Ansatz aufgenommen wurde.

Analog erlaubt die Codierung einer Profilkonstruktion (z. B. Codierung 4.1.1) in Kombination mit der Codierung einer Strategie und Idee zur Profilkonstruktion (z.B. S_3), eine differenzierte Aussage über die Experimentiersequenz. Die Lesart des Beispiels besagt, dass die Probanden ein Profil auf Basis von Messdaten durch Parameterkombination konstruiert haben (numerische Codierung) und zwar durch Vereinigung von Bedingungen zur Erzeugung von maximalen Spannungswerten mehrerer Variablen (symbolische Codierung).

Anwendung des Codiersystems. Das Codiersystem wird zur Auswertung der Beobachtungsprotokolle unter Einsatz von Codiervorschriften von drei geschulten Codierern und dem Autor der Studie angewendet. Jede Beobachtungsmitschrift wird nach folgendem Verfahren doppelt codiert: In einem ersten Schritt werden die Beobachtungsdaten von beiden Codierern unabhängig codiert. Anschließend folgt ein Abgleich der Ergebnisse mit Konsensbildung. Die Erstcodierung führt ein trainierter Codierer aus, die Zweitcodierung führt der Autor der Studie durch. Dies hat folgenden Grund: Nach einem Basistraining, in dem alle Codierer u. a. ausgewählte Beobachtungsprotokolle von Probandenpaaren der Kontroll- und Testgruppen ausgewertet haben, wurden die Codierer entweder auf die Auswertung der Beobachtungsmitschriften der Kontrollgruppe (KG) oder auf die Auswertung der Beobachtungsmitschriften der Testgruppen (TG 1 – 3) spezialisiert. Diesem Verfahren wurde unterstellt, dass Effekte, die spezifisch für die Kontrollgruppe oder spezifisch für die Testgruppen sind, erfolgreicher identifiziert werden können. Die Balancierung der Codierungen zwischen den Test- und der Kontrollgruppe stellt der Zweitcodierer her. Um Auswertereffekte wie Hypothesenbefangenheit o. ä. zu vermeiden, haben die geschulten Codierer keine Kenntnisse über die Forschungsfragen und über die Gütekriterien der Beobachter (vgl. Beobachterübereinstimmung). Zudem ist jedem Erstcodierer nur eine Teilmenge aller ausgewerteten Beobachtungsmitschriften bekannt.

Gütekriterien des Messinstruments. In der qualitativen Forschung wird eine Diskussion der Gütekriterien häufig unter der Terminologie „Kriterien der Validität“ geführt (Bortz & Döring, 2002, S. 261; Diekmann, 2007, S. 261). Diese Terminologie wird bei der Diskussion der Gütekriterien des qualitativen Messinstruments Beobachtungsbogen angewendet, indem auf eine formale Abgrenzung zwischen den zentralen Gütekriterien Objektivität, Reliabilität und Validität verzichtet wird. Nach Diekmann (ebd.) sollte ein qualitatives Messinstrument mindestens eines der Validitätskriterien erfüllen.

Mit den Kriterien zur Validierung des Messinstruments wird auf die Probleme und Fehlerquellen reagiert, die sich bei Beobachtungsstudien einstellen können. Die

Probleme können konzentriert in zwei Fragestellungen gefasst werden. Die erste Frage, mit der das Problem der Verzerrung durch selektive Wahrnehmung aufgegriffen wird, bezieht sich auf die Validität des Datenaufnahmeprozesses. Sie betrifft den Beobachter, der als Datenaufnehmer fungiert. Die zweite Frage, mit der das Problem der (Fehl-) Interpretation der beobachteten Experimentierhandlungen aufgegriffen wird, bezieht sich auf den Datenauswertungsprozess. Sie betrifft den Codierer, der die Beobachtungsdaten interpretiert. Die Fragestellungen sind der Literatur (Diekmann, 2007, S. 558; Bortz & Döring, 2002, S. 328 und S. 335) sinngemäß entnommen.

1. Bilden die Beobachtungsmitschriften die Experimentierhandlungen der Probandenpaare valide ab, oder sind sie durch den Datenaufnehmer z. B. durch Voreingenommenheit und Unaufmerksamkeit selektiv verzerrt oder verfälscht worden?
2. Ist die Interpretation der Experimentierhandlungen aus den Beobachtungsmitschriften valide? (interne Validität bzw. Gültigkeit von Interpretationen).

Die Dokumentation der Maßnahmen, mit denen mögliche Fehlerquellen der Beobachtung begegnet werden, orientiert sich an den beiden Fragestellungen. Die Dokumentation der Maßnahmen erfolgt in den Abschnitten Validitätskriterien der Beobachtung, Bestimmung der Beobachterübereinstimmung und Validitätskriterien der Auswertung.

Validitätskriterien der Datenaufnahme (Beobachtung). Der Beobachtungsbogen ist ein strukturiertes Messinstrument, dessen Einsatz durch detaillierte Instruktionen an die Beobachter reguliert wird. Durch die Kombination der Maßnahmen „Strukturierung des Messinstruments“ und „Beobachtungsvorschriften“ liegt die Beobachtung nicht im Ermessensspielraum des Datenaufnehmers. Diese Maßnahmen erhöhen die Durchführungsobjektivität und Zuverlässigkeit der Methode (Diekmann, 2007, S.567 ff.). Eingesetzt wird der Beobachtungsbogen von geschulten Beobachtern. Das Training dient der Validierung des Messinstruments. In Verbindung mit der Beobachterübereinstimmung ist die Beobachterschulung eine wichtige Voraussetzung für eine valide Interpretation des Beobachtungsmaterials (Bortz & Döring, 2002, S. 274 und S. 328).

Über die genannten Maßnahmen hinaus, wird die Validität der Beobachtungsmitschriften durch zusätzliche Kriterien gestützt. Um die Gefahr einer bevorzugten Aufnahme bestätigender Informationen (Bestätigungsbias) zu reduzieren, die eine Folge selektiver Wahrnehmung sind (Diekmann, 2007, S. 51 ff.), wird die Kontrolle und Systematik der Beobachtung erhöht. Dazu gehört die Kontrolle von Beobachtereffekten wie die Hypothesenbefangenheit, die randomisierte Zuweisung der Beobachter zu den Probandenpaaren, die Kreuzung mehrerer Datenerhebungsmethoden (Methodentriangulation) und, wenn die Freiwilligkeit der Probandenpaare

erfüllt ist, der punktuelle Einsatz des technischen Hilfsmittels Videografie (vgl. Kap. 4.4.1). Nicht befolgt wird die Empfehlung von Friedrich (1990, S. 287) zur Beobachtung eines Probandenpaares gleichzeitig zwei oder mehrere Beobachter einzusetzen. Denn wiederholte Erprobungen der Methode haben ergeben, dass Beobachterpaare während des Datenaufnahmeprozesses zu einer Konsensbildung neigen. Hier ist die Unabhängigkeit der Beobachter – eine Voraussetzung der Methode – nicht erfüllt worden.

Ermittlung der Beobachterübereinstimmung. Die Beobachterübereinstimmung wurde unter Einsatz einer videografischen Aufzeichnung von beobachtbaren Experimentiertätigkeiten und verbalen Argumentationen eines Probandenpaares ermittelt. Dazu mussten die 12 ausgebildeten Beobachter mit Hilfe des Beobachtungsbogens ein schriftliches Protokoll von den ausgewählten Experimentierprozessen des Probandenpaares erstellen und zwar bei einmaliger Vorführung (1. Kriterium) und ohne Unterbrechung (2. Kriterium) der Videosequenzen. Aufgrund der Kriterien werden an den Beobachter während der Ermittlung der Beobachterübereinstimmung höhere Anforderungen gestellt, als während der Aufnahme von Experimentiertätigkeiten in der Live-Beobachtung notwendig sind. Denn anders als in der Live-Beobachtung entfällt für den Beobachter während des Beobachterübereinstimmungsprozesses die Möglichkeit, an das beobachtete Probandenpaar Nachfragen zu stellen. Die Möglichkeit des Nachfragens bietet den Vorteil, um die von den Probandenpaaren gelieferten Informationen protokollarisch zu komplettieren. Das kann notwendig werden, wenn Probandenpaare Experimentierphasen mit alternierenden Informationsdichten gestalten, wie sie z. B. durch Überlagerung von beobachtbaren Experimentiertätigkeiten mit verbalen Argumentationen entstehen können. Es ist das Ziel, die Beobachter auf diese Fähigkeit zu testen, unter der Bedingung, dass die beiden Kriterien (einmalige Vorführung, ohne Unterbrechung) bei der Präsentation der videografischen Aufzeichnung eingehalten werden. Allein die Anzahl der verbalen Argumentationen (vgl. Anzahl der symbolischen Codierungen mit Pfeil in Abbildung 6) dokumentiert für die gewählte Experimentiersequenz, dass die von den Probanden transportierte und von den Beobachtern zu erfassende Informationsdichte sehr hoch ist. Dies war eins der Motive, die für die Auswahl der Videosequenz ausschlaggebend waren.

Zur Bestimmung der Beobachterübereinstimmung wurden auf Basis der Beobachtermitschriften individualisierte Beobachtungsprofile erstellt. Sie enthalten eine Sammlung von Ereignissen und Elementarereignissen, die unter Berücksichtigung der zeitlichen Reihenfolge aus dem Beobachtungsbogen übertragen wurden. Verglichen werden die Beobachterprofile der Datenaufnehmer mit einem Referenzbeobachter (Referenzprofil Video). Für ihn galten die Kriterien „einmalige Vorführung und „ohne Unterbrechung“ der Videosequenz nicht. Um den Einfluss eines Datenverlustes auf den Codierprozess zu kontrollieren, der auf den

Datenerhebungsprozess der Beobachter zurückzuführen ist, wird jedes Beobachterprofil unter Anwendung der Codierregeln in ein codiertes Beobachterprofil überführt. Die Prozessdiagramme der Abbildung 6 zeigen das uncodierte und codierte Profil des Referenzbeobachters für die ausgewählte Videosequenz. Zur detaillierten Analyse von Prozessdiagrammen sei auf das Kapitel 4.4.1 verwiesen.

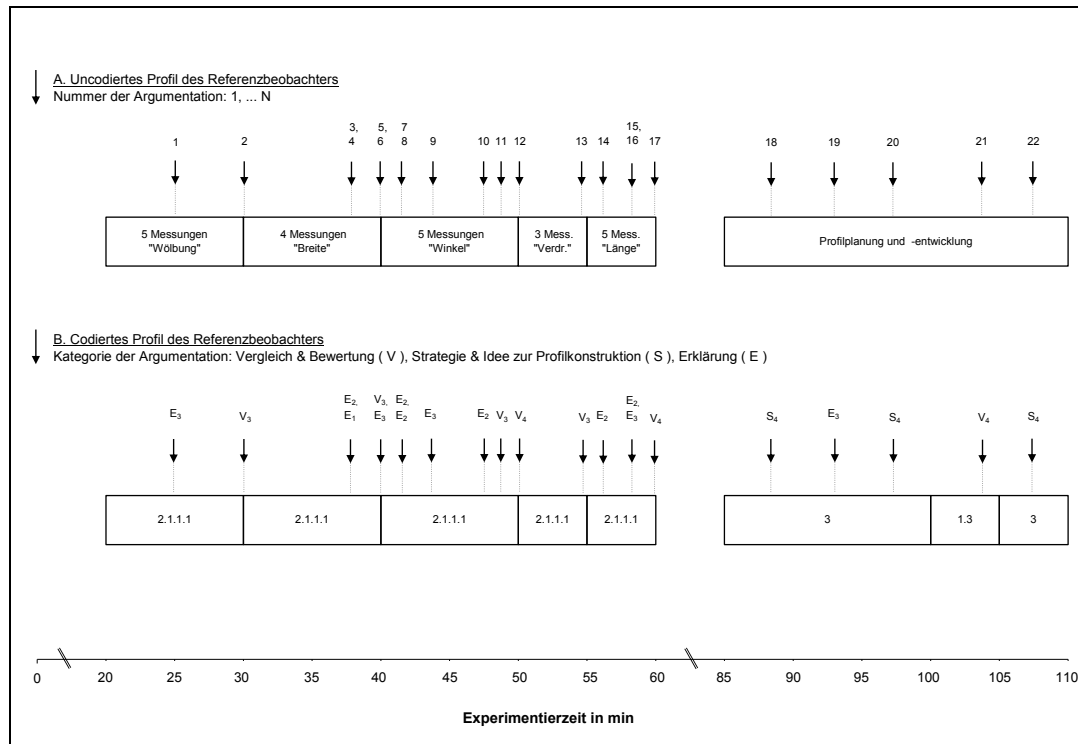


Abb. 6: Prozessdiagramme des Referenzbeobachters (Video) zur Ermittlung der Beobachterübereinstimmung

Der Vergleich der Beobachterprofile mit dem Referenzprofil ergibt für die Kategorien der beobachtbaren Experimentiertätigkeiten (numerische Codierungen) nur geringfügige Abweichungen. Sie bleiben für die Codierung folgenlos. Eine hohe Beobachterübereinstimmung kann auch für die verbalen Argumentationen ermittelt werden und zwar für die Dimension „zeitliche Anordnung der Beobachtungen“ und für die Dimension „Eindeutigkeit der Kategoriezuordnung der Beobachtungen“. Im Vergleich mit dem Referenzprofil zeigen sich jedoch Abweichungen bezüglich der Vollständigkeit (Anzahl) der erfassten verbalen Argumentationen. Dokumentiert werden die Ergebnisse mit der Tabelle 5. Sie bildet für alle geschulten Beobachter (Nr. 1 – 12) und für den Referenzbeobachter (Video) die Anzahl der erfassten verbalen Argumentationen ab. Aufgetragen sind die Nummern der Argumentationen, die die zeitliche Reihenfolge der Argumentationen widerspiegeln. Sie wurden aus den uncodierten, individualisierten Beobachtungsprofilen des Referenzbeobachters (Abbildung 6) und der 12 Datenaufnehmer ausgelesen.

Ermittlung der Beobachterübereinstimmung															
I	II	Nummer des Beobachters												Σ	
		0	1	2	3	4	5	6	7	8	9	10	11		12
E ₁	4	1	0	0	0	0	1	1	0	0	0	0	0	0	3
	E ₂ 3	1	0	1	0	0	1	1	0	1	1	0	0	0	6
E ₃	7	1	1	1	1	1	1	1	1	1	1	1	1	1	13
	8	1	0	1	1	0	1	0	1	1	0	1	0	1	8
	10	1	1	1	1	0	1	0	0	0	0	0	0	0	5
	14	1	1	1	1	1	1	1	1	1	1	1	1	1	13
	15	1	1	1	1	1	1	0	1	1	1	1	1	1	12
	1	1	1	1	1	1	1	1	0	1	1	1	1	1	12
	6	1	1	1	1	1	1	1	1	1	1	1	0	1	12
	9	1	1	1	0	1	1	1	0	1	1	0	1	1	10
	16	1	1	1	0	1	1	1	1	1	1	0	1	0	10
	19	1	1	0	1	0	1	0	1	1	1	1	0	0	8
V ₃	2	1	1	1	1	1	0	1	1	1	1	1	1	1	12
	5	1	1	1	1	1	0	1	1	1	0	1	1	0	10
	11	1	1	0	1	1	1	1	1	0	0	1	1	1	10
	13	1	1	0	1	1	0	1	1	0	1	1	0	0	8
V ₄	12	1	1	1	1	1	1	0	1	0	1	0	0	1	9
	17	1	1	1	1	1	1	1	1	1	0	1	1	1	12
	21	1	1	1	1	1	0	1	1	1	0	0	0	0	8
S ₄	18	1	1	1	1	1	1	1	1	1	1	1	1	1	13
	20	1	1	1	1	1	1	1	1	1	1	1	1	1	13
	22	1	0	1	1	1	0	0	0	0	1	0	1	0	6
Σ		22	18	18	18	17	17	16	16	16	15	14	13	13	
Legende: I / II Subkategorie der verbalen Argumentation im codierten (I) / uncodierten (II) Beobachterprofil (vgl. Abbildung 6) 1 / 0 Verbale Argumentation vom Datenaufnehmer in Beobachtungsmitschrift protokolliert (1) / nicht protokolliert (0) Nr. 0 Referenzbeobachter Video Kategorien der verbalen Argumentationen: Erklärung (E), Vergleich und Bewertung einer Messfolge (V), Strategie & Idee zur Profilkonstruktion (S)															

Tab. 5: Anzahl der erfassten verbalen Argumentationen zur Ermittlung der Beobachterübereinstimmung

Die Abweichungen bzgl. der Anzahl der erfassten verbalen Argumentationen können auf die technische Überlegenheit des Referenzbeobachters zurückgeführt werden. Trotzdem bleibt festzuhalten, dass die Kategorien der verbalen Argumentationen in einer ausreichenden Anzahl durch die Beobachter abgebildet

werden, so dass die Anforderungen an einen kategoriengeleiteten Messprozess unter Einsatz des Beobachtungsbogens erfüllt werden. Für die Schlussfolgerung war nicht nur die Anzahl der erfassten verbalen Argumentationen, sondern auch ihre Verteilung über die Subkategorien ausschlaggebend. Wie der Tabelle 5 zu entnehmen ist, wird jede Subkategorie von den Beobachtern (Nr. 1 – 12) mehrfach erfasst. Davon ausgenommen ist die Subkategorie E_1 , die in der Videosequenz nur als einmaliges Ereignis registriert werden konnte. Es stellt sich die Frage, welche Relevanz ein singuläres Ereignis für die Ermittlung der Beobachterübereinstimmung hat.

Validitätskriterien der Datenauswertung (Codierung). Für die Auswertung von qualitativen Daten ist die konsensuelle Validierung das wichtigste Gütekriterium (Bortz & Döring, 2002, S. 328). Diese Methode der interpersonalen Konsensbildung wurde für die Auswertung der Beobachtungsmitschriften eingesetzt (vgl. *Anwendung des Codiersystems*). Fehlerquellen, die eine Codierung und Konsensbildung beeinträchtigen, sind erstens ein unsystematisches Codiersystem und zweitens unklare Codierkategorien (Friedrichs, 1990, S. 288). Der ersten Fehlerquelle wird mit einem strukturierten Codiersystem begegnet. Zudem werden Experimentierhandlungen, die einen Interpretationsspielraum erlauben, durch Codiervorschriften kontrolliert und auf ihre Relevanz für die Eingruppierung in die Niveaustufen des Experimentierens geprüft. Als Ergebnis der Prüfung konnte festgestellt werden, dass die Kriterien zur Eingruppierung in die Niveaustufen einen Toleranzbereich zulassen, mit dem Interpretationsunschärfen aufgefangen bzw. kompensiert werden können. Als ein Beispiel sei die Aufnahme einer konfundierenden Messfolge genannt, die auch als eine Aneinanderreihung konfundierender Einzelmessungen gedeutet werden könnte. Aber weder die eine noch die andere Interpretationsvariante hätte eine Konsequenz für die Eingruppierung in die Niveaustufen des Experimentierens. Beide Deutungen der Experimentiersequenz würden zur Eingruppierung in dieselbe Niveaustufe führen (vgl. Kap. 4.4.2, Niveaustufe 1).

Der zweiten Fehlerquelle wurde mit einer präzisen Beschreibung der Codierkategorien (Subkategorien) begegnet. Angestrebt wurde das Ziel, die Kategorien (Subkategorien) disjunkt und erschöpfend zu formulieren. Zu erkennen ist das beispielsweise an der Codierung von beobachtbaren Experimentiertätigkeiten, denen eine Codierung von komplementären, beobachtbaren Experimentiertätigkeiten gegenübergestellt wird (Aufnahme einer Messfolge/ Einzelmessung, nicht konfundierend/ konfundierend, Profilkonstruktion messdatenbasiert/ nicht messdatenbasiert). Das trifft auch für die Codierung der Kategorie Profil konstruieren mit den Subkategorien Profilkonstruktion durch Parameterkombination und Profilkonstruktion durch Kopie einer Profilverlage zu. Die Codiervorschrift reguliert, dass alles, was nicht Kopie einer Profilverlage ist, als Profilkonstruktion

durch Parameterkombination codiert wird. Diese Lösung erlaubt, dass jede Variante möglicher Profilkonstruktionen eindeutig einer Kategorie zugeordnet werden kann. Die Kategorie ist disjunkt und erschöpfend. Die genannten Kriterien waren auch bei der Formulierung der Kategorien der verbalen Argumentationen maßgebend, die im Gegensatz zu den beobachtbaren Experimentiertätigkeiten mit polytomer Ausprägung vorliegen. In Zweifelsfällen hat der Codierer die Möglichkeit, Experimentierhandlungen mit der Kategorie „Sonstiges“ (Code 0) zu verschlüsseln. Die Kategorie wird nicht systematisch ausgeschlossen, weil für das empirisch erstellte Codiersystem nur Experimentierhandlungen zu Kategorien zusammengefasst wurden, die aus den Beobachtungsmitschriften wiederholt ausgelesen werden konnten.

Angemerkt sei, dass auf die Ermittlung der Intercoderreliabilität verzichtet wurde. Zwei Argumente haben zu der Entscheidung geführt. Das erste Argument betrifft die Verhältnismäßigkeit der Methoden, die an den Zielen einer hypothesengenerierenden Studie ausgerichtet wurden. Das führte zu der Entscheidung, dass entweder die Beobachter- oder die Intercoderreliabilität bestimmt wird. Die Entscheidung zugunsten der Ermittlung der Beobachterreliabilität wird durch die Wahl der Beobachtungsmethode und den Einsatz des Messinstruments Beobachtungsbogen begründet (2. Argument). Denn die Datenaufnahme durch die Person des Beobachters wird im Vergleich zur Datenauswertung durch die Person des Codierers als der sensiblere und fehleranfälligere Prozess eingeschätzt. Begründet wird die Einschätzung mit den Anforderungen, die an den Protokollanten gestellt werden: Er muss die Beobachtungsdaten in zeitlich korrekter Abfolge zuverlässig selektieren und den Anspruch der Datenvollständigkeit erfüllen. Dem Codierer von Beobachtungsmitschriften verschafft das den Vorteil, dass ihm nur codierrelevantes Datenmaterial vorliegt.

4.3 Ergebnisse des Fach- & NAW-Tests

Mit dem Einsatz des Fach- & NAW-Tests werden zwei Ziele verfolgt. Erstens werden die Experimentalgruppen der Teilstichprobe auf Basis der Ergebnisse des NAW-Vortests zum Messzeitpunkt 1 unter Kontrolle des Fachwissens (Fach-Test) parallelisiert. Zweitens werden die allgemeinen Fähigkeiten im Umgang mit Variablen über alle drei Messzeitpunkte auf Grundlage der Ergebnisse des NAW-Tests erfasst. Zur Auswertung der Messdaten werden t-Tests und einfaktorielle Varianzanalysen benutzt. Zu den gemeinsamen Voraussetzungen beider statistischer Verfahren gehört u. a., dass die Messdaten normalverteilt sind. Der erste Abschnitt widmet sich deshalb einer Prüfung der Voraussetzungen zum Einsatz der statistischen Auswertungsverfahren.

Prüfung der Voraussetzungen zum Einsatz der statistischen Auswertungsverfahren. Die Messdaten werden mit dem Kolmogorow-Smirnow-Test (KS-Test) auf Normalverteilung geprüft. Der Test misst die Anpassung einer empirischen Verteilungsfunktion an eine kontinuierliche, theoretische Verteilungsfunktion. Der Anpassungstest darf auch bei diskreten Verteilungen angewandt werden (Diehl & Arbinger, 2001, S. 514). Die Nullhypothese behauptet, dass zwischen der empirischen und der theoretischen Verteilung kein Unterschied existiert, die Alternativhypothese behauptet das Gegenteil. Die Entscheidung über die Hypothesen wird durch eine Teststatistik mit der Prüfgröße d herbeigeführt. Die Prüfgröße d der Kolmogorow-Smirnow-Differenzstatistik ist das Supremum des Ordinatsabstandes der theoretischen und empirischen Verteilungsfunktion. Wenn d einen definierten, kritischen Grenzwert d_{krit} unterschreitet, kann eine signifikante Abweichung der empirischen Verteilung von der Normalverteilung nicht angenommen werden. In der vorliegenden Studie werden alle kritischen Grenzwerte des KS-Tests für den zweiseitigen Fall mit $\alpha = 0,05$ gewählt. Bei der Durchführung des Tests muss beachtet werden, dass mit der Wahl des Auswertungsverfahrens die Grundgesamtheit der Probanden variiert, woraus eine Verschiebung der kritischen Grenzen resultiert.

Die einfaktorielle Varianzanalyse ohne Messwiederholung wird für die Parallelisierung der Teilstichprobe eingesetzt. Zur Teilstichprobe zählen 82 Probanden. Der korrespondierende kritische Grenzwert wird in der Literatur mit $d_{\text{krit}} = 0,15$ beziffert (Diehl & Arbinger, S. 765). Der kritische Wert wird zum Messzeitpunkt 1 von dem empirischen d -Wert des Fach-Tests mit $d = 0,10$ und von dem empirischen d -Wert des NAW-Tests mit $d = 0,12$ unterschritten, so dass die Voraussetzungen für den Einsatz der einfaktoriellen Varianzanalyse ohne Messwiederholung erfüllt sind. Die Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen werden über die Messzeitpunkte 1 – 3 für jede Experimentalgruppe separat durch t -Tests für abhängige Stichproben (Teilstichproben) dokumentiert. In der Kontrollgruppe, der Testgruppe 1 und der Testgruppe 3 befinden sich je 24 Probanden, in der Testgruppe 2 befinden sich 10 Probanden. Die korrespondierenden kritischen Grenzwerte werden mit $d_{\text{krit}} = 0,27$ ($N = 24$ Probanden) und mit $d_{\text{krit}} = 0,41$ ($N = 10$ Probanden) benannt. Bei abhängigen Stichproben muss die Voraussetzung erfüllt sein, dass bei kleinen Stichprobenumfängen mit $n < 30$ und $n = \text{Anzahl der Messwertpaare}$, die Differenzen der Messwertpaare normalverteilt sind (Bortz, 2005, S. 144). Aus diesem Grund werden für den NAW-Test die Differenzen der Messwertpaare der Messzeitpunkte 1 und 2 (Synonyme Schreibweise: 1/ 2), der Messzeitpunkte 2 und 3 (Synonyme Schreibweise: 2/ 3) sowie der Messzeitpunkte 1 und 3 (Synonyme Schreibweise: 1/ 3) einer Normalverteilungsprüfung unterzogen.

KS-Test auf Normalverteilung für abhängige Stichproben mit Differenzen der Messwertpaare im NAW-Test zu den Messzeitpunkten					
Gruppe	N	d_{krit}	$d_{1/2}$	$d_{2/3}$	$d_{1/3}$
KG	24	0,27	0,25	0,13	0,18
TG 1	24	0,27	0,23	0,16	0,17
TG 2	10	0,41	0,28	0,34	0,22
TG 3	24	0,27	0,21	0,17	0,13
Legende: KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3) N Anzahl der Probanden d Supremum des Ordinatsabstandes der theoretischen und empirischen Verteilungsfunktion d_{krit} kritischer d-Wert für den zweiseitigen Fall mit $\alpha = 0,05$ $d_{1/2}$, empirische d-Werte für Differenzen der Messwertpaare im NAW-Test $d_{2/3}$, $d_{1/3}$ zu den Messzeitpunkten 1 und 2, 2 und 3 sowie 1 und 3					

Tab. 6: Kolmogorow-Smirnow-Anpassungstests (KS-Test) auf Normalverteilung für abhängige Stichproben mit Differenzen der Messwertpaare im NAW-Test zu den Messzeitpunkten.

Fettdruck: signifikante Abweichung mit $d > d_{krit}$

Wie der Tabelle 6 entnommen werden kann, unterschreiten alle empirischen d-Werte die korrespondierenden kritischen Grenzwerte d_{krit} , so dass die Voraussetzungen zum Einsatz von t-Tests für abhängige Stichproben (Teilstichproben) erfüllt sind.

Eine weitere Voraussetzung für die Durchführung von Varianzanalysen und t-Tests ist die Gleichheit der Varianzen in den zu untersuchenden Gruppen. Geprüft wird, ob sich die Varianzen von zwei oder mehr Stichproben (Teilstichproben) statistisch signifikant unterscheiden. Zum Messzeitpunkt 1 ist die Varianzhomogenität für die unabhängigen Experimentalgruppen (KG, TG 1 – 3) sowohl für die Dimension Naturwissenschaftliche Arbeitsweisen als auch für die Dimension Fachwissen des Fach- & NAW-Tests erfüllt. Sie wurde mit dem Levene-Test bestimmt, der für die Dimension Naturwissenschaftliche Arbeitsweisen zu dem Ergebnis $F_{NAW}(3, 78) = 0,12$ mit $p = 0,946$ und für die Dimension Fachwissen zu dem Ergebnis $F_{Fach}(3, 78) = 1,95$ mit $p = 0,128$ führt.

Im Kontrast zum Varianzhomogenitätstest für unabhängige Gruppen ist eine präzise Antwort auf die Frage, ob zwischen den Varianzen von mehr als zwei abhängigen Stichproben statistisch signifikante Unterschiede bestehen, wegen eines unvollkommenen Testverfahrens nicht immer möglich. Als ein Nachteil wird beispielsweise benannt, dass das Testverfahren nur für große Stichproben ($N \geq 40$) geeignet ist. Der Nachteil kann für die vorliegende Studie geltend gemacht werden. Notwendig wäre ein Testverfahren, das die Varianzhomogenität für drei abhängige Stichproben mit einer Stichprobengröße von $N < 40$ prüft.

Alternativ wird in der Literatur der Rückgriff auf ein Verfahren vorgeschlagen, dass die Varianzhomogenität für genau zwei abhängige Stichproben testet. Es soll mehrfach angewendet werden (Diehl & Arbinger, 2001, S. 360 ff.). Da die Alternative nur eine von mehreren nachrangigen Kompromisslösungen ist, wird auf eine Testung der Varianzhomogenität verzichtet.

Parallelisierung der Teilstichprobe. Die Experimentalgruppen der Teilstichprobe werden nach den Ergebnissen der Fach- & NAW-Tests parallelisiert. Für die Parallelisierung sind die Messergebnisse der Dimension Naturwissenschaftliche Arbeitsweisen (NAW-Test) des Messinstruments zum Messzeitpunkt 1 ausschlaggebend (vgl. Abbildung Anhang A 4). Zudem wird das Fachwissen der Experimentalgruppen nach den Ergebnissen des Fach- & NAW-Tests kontrolliert. Zur Kontrolle des Vorwissens sind die Messergebnisse der Dimension Fachwissen (Fach-Test) zum Messzeitpunkt 1 ausschlaggebend. Bei beiden Vergleichen handelt es sich um Datenanalysen für unabhängige Stichproben. Die Varianzanalysen führen weder für die Dimension Fachwissen (Fach) noch für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) zu einem signifikanten Ergebnis.

Die Resultate deuten darauf hin, dass die Probanden der Experimentalgruppen über dieselben fachinhaltlichen und fachmethodischen Voraussetzungen verfügen und dass die Parallelisierung der Experimentalgruppen erfolgreich vollzogen wurde. Die einfaktorielle Varianzanalyse ohne Messwiederholung ergibt für die Kontrolle des Fachwissens $F_{\text{Fach}}(3, 78) = 1,63$ mit $p = 0,190$ und für die Parallelisierung der Teilstichprobe $F_{\text{NAW}}(3, 78) = 0,06$ mit $p = 0,981$.

Über die Varianzanalyse hinaus werden die Gruppenvergleiche für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) zum Messzeitpunkt 1 grafisch durch die Abbildung 7 ergänzt. Aus den Vergleichen der Mittelwerte und den Vergleichen der Konfidenzintervalle für die Mittelwerte wird ersichtlich, dass für die Experimentalgruppen KG, TG 1 und TG 3 die Konfidenzintervalle annähernd gleich breit sind. Dagegen übersteigt die Konfidenzintervallbreite der Testgruppe 2 die Konfidenzintervallbreite der anderen Experimentalgruppen, was auf eine inhomogenere Gruppenzusammensetzung hindeutet. Die Breite eines Konfidenzintervalls hängt vom Stichprobenumfang und von der Streuung der Datenwerte ab. Aus der

Tabelle 7 kann für die Testgruppe 2 zum Messzeitpunkt 1 entnommen werden, dass bei einem kleinen Stichprobenumfang eine relativ hohe Streuung vorliegt.

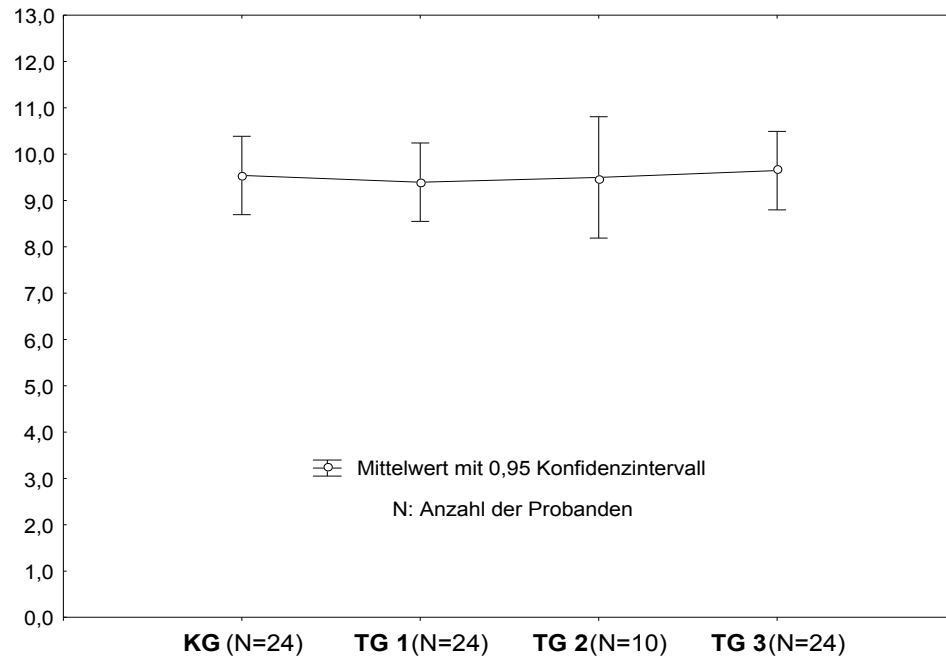


Abb. 7: Gruppenvergleiche für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests zum Messzeitpunkt 1 (Vortest). Wertebereich 0 – 13 Punkte

Änderung der allgemeinen Fähigkeiten im Umgang mit Variablen. Um die Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen zu erfassen, werden die Testergebnisse der Probanden der Teilstichprobe gruppenweise zu den Messzeitpunkten 1 – 3 verglichen (N: Anzahl der Probanden; df = N-1: Anzahl der Freiheitsgrade). Zum Messzeitpunkt 1 wird der Vortest, zum Messzeitpunkt 2 der Nachtest und zum Messzeitpunkt 3 der Verzögerte Nachtest durchgeführt. Die Testergebnisse werden mit t-Tests ausgewertet, weil die Änderungen von Messzeitpunkt zu Messzeitpunkt akzentuiert werden sollen. Für die statistischen Berechnungen wird der t-Test für abhängige Stichproben (Teilstichproben) eingesetzt, weil die Unabhängigkeit der Messdaten bei wiederholten Testungen nicht erfüllt ist. Infolgedessen muss auch die Abhängigkeit der Stichprobe (Teilstichprobe) bei Angabe eines Effektstärkenmaßes berücksichtigt werden. Das Effektstärkenmaß d' – eine Modifikation des Effektstärkenmaßes d nach Cohen für unabhängige Stichproben (Teilstichproben) – erfüllt die Voraussetzungen (Bortz & Döring, 2002, S. 606 ff.). Es wird für die signifikanten Effekte in der Tabelle 7 verwendet.

t-Tests auf Unterschiede zwischen den Messzeitpunkten (MZ) der Experimentalgruppen KG & TG 1 – 3 für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW)						
MZ		MW	SD	t-Wert	p-Wert	d`
KG	df = 23					
1		9,54	2,22	0,00	1,000	n. s.
2		9,54	2,44			
2		9,54	2,44	1,05	0,303	n. s.
3		9,89	2,46			
TG 1	df = 23					
1		9,40	1,94	3,07	0,005	0,63
2		10,10	1,56			
2		10,10	1,56	1,74	0,096	n. s.
3		10,56	1,20			
TG 2	df = 9					
1		9,50	2,20	0,40	0,696	n. s.
2		9,70	2,62			
2		9,70	2,62	2,69	0,025	0,85
3		10,40	2,31			
TG 3	df = 23					
1		9,65	2,02	3,68	0,001	0,75
2		10,52	1,53			
2		10,52	1,53	1,95	0,064	n. s.
3		9,85	2,53			
Legende:						
KG	Kontrollgruppe					
TG	Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3)					
t	t-Wert für abhängige Stichproben (Teilstichproben)					
d`	Effektstärke für abhängige Stichproben (Teilstichproben)					
MZ	Messzeitpunkt: Vortest (1), Nachtest (2), Verzögerter Nachtest (3)					

Tab. 7: t-Tests auf Unterschiede zwischen den Messzeitpunkten 1 und 2 sowie 2 und 3 der Kontroll- und Testgruppen für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests. **Fettdruck:** signifikante Effekte für $p < 0,05$

Zur Klassifikation des Effektstärkenmaßes d' werden in der Literatur folgende Grenzen genannt: Ein kleiner Effekt liegt für $0,2 \leq d' < 0,5$ vor, ein mittlerer Effekt liegt für $0,5 \leq d' < 0,8$ vor und ein großer Effekt liegt für $0,8 \leq d'$ vor (Bortz & Döring, 2002, S. 604). An den Grenzen ist zu erkennen, dass die Effektstärken-

einteilung für d und d' übereinstimmen, denn nach Bortz & Döring (2002, S. 605) gilt die Effektstärkenklassifikation für den t-Test mit unabhängigen Stichproben auch für den t-Test mit abhängigen Stichproben. Relevant ist zudem, dass in das modifizierte Effektstärkenmaß d' der Korrelationskoeffizient r einer Produkt-Moment-Korrelation einfließt, der in den ausgewiesenen Ergebnissen der Tabelle 7 für die Produkt-Moment-Korrelation nach Pearson berechnet wurde. Die Korrelationskoeffizienten r können der Tabelle 8a entnommen werden.

Wie der Tabelle 7 zu entnehmen ist, führen die t-Tests für die Dimension Naturwissenschaftliche Arbeitsweisen des Fach- & NAW-Tests nicht immer zu signifikanten Unterschieden, wenn die Messzeitpunkte 1/ 2 und 2/ 3 paarweise für jede Experimentalgruppe auf einem Signifikanzniveau von 5 % verglichen werden. Während die Änderungen vom Vor- zum Nachtest und vom Nachtest zum Verzögerten Nachtest für die Kontrollgruppe nicht signifikant sind, ist bei den Testgruppen Reproduktion (TG 1) und Transfer (TG 3) zwischen dem Vor- und Nachtest ein signifikanter Anstieg mit mittlerer Effektstärke festzustellen. Zudem stimmen die Testgruppen 1 und 3 darin überein, dass beim Vergleich des Nachtests mit dem Verzögerten Nachtest keine signifikanten Änderungen zu verzeichnen sind, obwohl ein zeitlicher Abstand von 12 – 13 Wochen zwischen den Messzeitpunkten liegt. Die Testgruppe Reorganisation (TG 2) hebt sich von den Testgruppen 1 und 3 ab, weil die Änderungen zwischen dem Vor- und Nachtest nicht signifikant sind und die Testergebnisse zwischen dem Nach- und dem Verzögerten Nachtest signifikant mit hoher Effektstärke ansteigen. Die Testgruppe 2 weist sich gegenüber den weiteren Experimentalgruppen mit $N = 10$ Probanden durch eine erheblich geringere Probandenanzahl aus.

Die Ergebnisse der t-Tests erlauben für die Experimentalgruppen eine interpretierende Diskussion über die Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen, die für die Testgruppen mit Berücksichtigung des Einflusses der vorbereitenden Instruktion und für die Kontrollgruppe mit Berücksichtigung des fehlenden Einflusses der vorbereitenden Instruktion geführt werden muss.

Für die Kontrollgruppe führen die fast übereinstimmenden statistischen Kennwerte im Vor- und Nachtest zu einem p von 1,000. Der Wert besagt, dass die statistische Datenanalyse im vorliegenden Fall keine signifikante Änderung der allgemeinen Fähigkeiten im Umgang mit Variablen ergeben hat. Die Ursache für die ausbleibende signifikante Änderung kann auf die fehlende vorbereitende Instruktion zurückgeführt werden, weil die Kontrollgruppe ohne Unterschiede zu den Testgruppen an der offenen Experimentieraufgabe partizipiert. Zudem besagt das Resultat der Kontrollgruppe, dass es den Probanden nicht gelingt, nur durch die Bearbeitung einer kontextspezifischen offenen Experimentieraufgabe zum Thema „Konstruktion von Rotorprofilen mit maximaler Drehzahl für das Modell eines Windrades“ (vgl. Kap. 2.3), ihre allgemeinen Fähigkeiten im Umgang mit Variablen

signifikant zu verbessern. Weiterhin ist nicht zu vernachlässigen, dass es denselben Probanden nach Abschluss der Erhebungseinheit aus Vortest, offener Experimentieraufgabe und Nachtest nicht gelingt, ihre allgemeinen Fähigkeiten im Umgang mit Variablen durch die Teilnahme am schulischen Physikunterricht zu ändern, der über einen Zeitraum von drei Monaten zwischen dem Nach- und dem Verzögerten Nachtest von den Fachlehrern erteilt wurde. In dem Zeitintervall standen viele der Probanden nicht nur unter dem Einfluss des schulischen Physikunterrichts, sondern auch unter dem Einfluss weiterer naturwissenschaftlicher Unterrichtsfächer (vgl. Kap. 4.2.1). Das Ergebnis kann aus dem nicht signifikanten Unterschied zwischen Nach- und Verzögertem Nachtest gefolgert werden und erlaubt die Interpretation, dass ein explizites Unterrichten von naturwissenschaftlichen Arbeitsweisen wie den Umgang mit Variablen nicht zu den praktizierten Unterrichtsmethoden der Fachlehrer zählte. Bewusst bezieht sich die Interpretation auf die fachspezifische Unterrichtsmethodik, weil die curricularen Vorgaben in Physik (NRW) für die Sekundarstufe I und für die Jahrgangsstufe 11 (G 9) eine Verschränkung von fachspezifischen Inhalten mit fachspezifischen Methoden wie der Variablenkontrollstrategie aufweisen. Die Vermutung ist, dass die fachspezifischen Methoden im Physikunterricht implizit vermittelt wurden und als Notwendigkeit fungierten, um die fachspezifischen Ziele zu erreichen. Im Gegensatz dazu steht bei einer expliziten Vermittlungsmethodik der Erwerb von naturwissenschaftlichen Arbeitsweisen im Fokus. Am Beispiel des Umgangs mit Variablen wurde nach dem didaktischen Ansatz der vorbereitenden Instruktion eine explizite Vermittlung naturwissenschaftlicher Arbeitsweisen erprobt. Im Folgenden wird die Wirkung der Maßnahme interpretiert. Die Deutung der Ergebnisse wird vor allem von den Ergebnissen der Testgruppen 1 und 3 getragen, weil die Zusammensetzung der Testgruppe 2 nicht passgenau mit der Zusammensetzung der anderen Experimentalgruppen übereinstimmt.

Für die Testgruppen 1 und 3 ist der signifikante Unterschied zwischen den Messzeitpunkten 1 und 2 als Steigerung der allgemeinen Fähigkeiten im Umgang mit Variablen zu deuten, den die einmalige vorbereitende Instruktion auf Grundlage einer expliziten Vermittlungsmethodik im zeitlichen Umfang einer Schulstunde (45 Minuten) bewirkt. Ein Vergleich der Ergebnisse der beiden Testgruppen mit den Ergebnissen der Kontrollgruppe lässt aufgrund der gelungenen zufallsverknüpften Parallelisierung die Folgerung zu, dass als Ursache für die Änderung der allgemeinen Fähigkeiten im Umgang mit Variablen erstens die Bearbeitung der offenen Experimentieraufgabe und zweitens die Kombination aus vorbereitender Instruktion und Bearbeitung der offenen Experimentieraufgabe ausgeschlossen werden kann. Vielmehr ist die Ursache allein in der vorbereitenden Instruktion zu suchen.

Weiterhin bemerkenswert ist, dass für die Testgruppen 1 und 3 keine signifikanten Änderungen vom Nach- zum Verzögerten Nachtest festzustellen sind. Das Ergebnis ist ein Indikator für eine längerfristige Wirkung der vorbereitenden Instruktion. Die Annahme, dass der vom Kurslehrer zwischen dem Nach- und Verzögerten Nachtest erteilte und nicht zu kontrollierende, schulische Physikunterricht als einflussreiche Störvariable agiert (vgl. Kap. 4.2.1), die die Aussagekraft über eine anhaltende Wirkung der vorbereitenden Instruktion verzerrt, kann auf Basis der vorliegenden Ergebnisse nicht aufrechterhalten werden. Denn von einem wirkungsvollen Physikunterricht, in dem naturwissenschaftliche Arbeitsweisen fokussiert und thematisiert werden, würden unter Einsatz desselben Testinstruments vom Messzeitpunkt 2 zum Messzeitpunkt 3 signifikante Zuwächse erwartet werden. In Analogie zur Kontrollgruppe wird der nicht signifikante Zuwachs der beiden Testgruppen auf eine implizite Vermittlungsmethodik von naturwissenschaftlichen Arbeitsweisen im Physikunterricht zurückgeführt. Die Prägnanz der Aussage, die aus den nicht signifikanten Testergebnissen abgeleitet wird, erhöht sich durch die Tatsache, dass die Probanden der drei Experimentalgruppen (KG, TG 1, TG 3) nach den Kriterien der Gruppenzusammensetzung von sechs verschiedenen Fachlehrern unterrichtet wurden. Für die Testgruppe Reorganisation (TG 2) konnten die Kriterien des Gruppenmatchings, nachdem die Probanden einer Experimentalgruppe aus zwei verschiedenen Physikkursen stammen sollen, nicht eingehalten werden. Das hat zur Konsequenz, dass die Ergebnisse der Testgruppe 2 von klumpenspezifischen Effekten verzerrt sein könnten. Der signifikante Anstieg der Testergebnisse vom Nach- zum Verzögerten Nachtest beruht vermutlich auf dem Einfluss dieser Störvariablen. Zusätzlich wird die Vermutung durch die ausgeprägte Effektstärke gestützt. Die Probanden und Kurslehrer waren sowohl für die Forschungsfragen als auch für die Testinstrumente blind. Trotzdem ist nicht ausgeschlossen, dass zum Messzeitpunkt 3 von den Probanden ein wiederholter Einsatz desselben Testinstruments vermutet worden war und infolge dessen eine gezielte, ergebnisverzerrende, nicht zu kontrollierende Testvorbereitung stattgefunden hat. Wenn die Vermutung stimmig ist, dann gibt das Testergebnis einen klumpenspezifischen Übungs- bzw. Trainingseffekt wieder.

Entgegen des signifikanten Zuwachses vom Messzeitpunkt 2 zum Messzeitpunkt 3 kann der nicht signifikante Zuwachs vom Vor- zum Nachtest der Testgruppe 2 vermutlich nicht auf einen klassen- bzw. klumpenspezifischen Effekt zurückgeführt werden, obwohl gegenüber den anderen Experimentalgruppen der Teilstichprobe (KG, TG 1, TG 3) eine Unterschreitung der mittleren Testbearbeitungszeit im Posttest von bis zu fünf Minuten und gegenüber der gesamten Teilstichprobe eine Unterschreitung der mittleren Testbearbeitungszeit von drei Minuten vorliegt. Die mittlere Testbearbeitungszeit der Teilstichprobe ($N = 82$) kann im Posttest mit 19 Minuten beziffert werden. Mit dem Zeitmaß können die Abweichungen der

Ergebnisse nicht erklärt werden, weil mit dem wiederholten Einsatz desselben Messinstruments für alle Experimentalgruppen vom Vor- zum Nachtest ungefähr eine Halbierung der mittleren Testbearbeitungszeit einhergeht. Demgegenüber kann als Ursache dafür, dass die vorbereitende Instruktion für die Testgruppe 2 keinen signifikanten Zuwachs der allgemeinen Fähigkeiten im Umgang mit Variablen bewirkt, die inhomogenere Verteilung der Testergebnisse der Probanden über den Punktebereich des Vortests geltend gemacht werden. Das hat zur Konsequenz, dass die Testgruppe Reorganisation (TG 2) weniger gut parallelisiert worden ist als die weiteren Experimentalgruppen (KG, TG 1, TG 3). Die Unterschiede der Testergebnisse im Vortest werden z. B. durch einen Vergleich der Konfidenzintervalle in Abbildung 7 oder im Kapitel 4.5 sichtbar. Die inhomogene Verteilung konnte nicht aufgelöst werden, weil zur Datenauswertung nur die Ergebnisse von 10 Probanden vorlagen.

Kritische Bewertung der statistischen Auswertungsverfahren. Die Ergebnisse des Fach- & NAW-Tests werden vor der Durchführung der vorbereitenden Instruktion zum Messzeitpunkt 1 für die vier voneinander unabhängigen Experimentalgruppen der Teilstichprobe (KG, TG 1 – 3) mit Hilfe einer einfaktoriellen Varianzanalyse (ANOVA) ausgewertet. Mit einer ergänzenden statistischen Analyse auf der Basis von mehreren t-Tests für unabhängige Stichproben (Teilstichproben) könnten zudem gruppenspezifische Unterschiede des Fachwissens und gruppenspezifische Unterschiede für die Dimension Naturwissenschaftliche Arbeitsweise stärker herausgestellt werden. Weil die Durchführung von mehreren t-Tests an denselben Messdaten erstens zu einer nicht akzeptablen α -Fehler-Kumulierung und zweitens zu einer Verringerung der Teststärke führen würde, wurde auf den zusätzlichen Einsatz von t-Tests für unabhängige Stichproben verzichtet. Die Teststärke, die sich proportional zur Anzahl der Probanden verhält, verringert sich mit dem Wechsel der Auswertungsverfahren, da die Stichprobengröße jeder Experimentalgruppe mit $N \leq 24$ Probanden kleiner ist als die Stichprobengröße der gesamten Teilstichprobe ($N = 82$).

Entgegen der vorausgehenden Argumentation werden die Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen gruppenspezifisch mit t-Tests für abhängige Stichproben (Teilstichproben) ausgewertet, indem die Ergebnisse des NAW-Tests zu den Messzeitpunkten 1 – 3 verglichen werden (Tabellen 8a und 8b). Trotz der genannten Vorteile wird auf eine Auswertung durch Varianzanalyse verzichtet, weil im Fall von Messwiederholungen die Abhängigkeit der Stichproben berücksichtigt werden muss.

Korrelationskoeffizienten r zwischen den Messzeitpunkten 1 – 3 der Experimentalgruppen für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests					
$r_{KG, N=24}$	2	3	$r_{TG 1, N=24}$	2	3
1	0,82	0,58	1	0,81	0,60
2	–	0,77	2	–	0,59
$r_{TG 2, N=10}$	2	3	$r_{TG 3, N=24}$	2	3
1	0,80	0,85	1	0,82	0,68
2	–	0,95	2	–	0,77
Legende: r Korrelationskoeffizient nach Pearson KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3) 1 – 3 Messzeitpunkt: Vortest (1), Nachtest (2), Verzögerter Nachtest (3) N Anzahl der Probanden					

Tab. 8a: Korrelationskoeffizienten r nach Pearson zwischen den Messzeitpunkten 1 – 3 der Experimentalgruppen für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests.

Signifikante Korrelationen zwischen den Messzeitpunkten 1 – 3 der Experimentalgruppen für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests					
$p_{KG, N=24}$	2	3	$p_{TG 1, N=24}$	2	3
1	< 0,001	0,003	1	< 0,001	0,002
2	–	< 0,001	2	–	0,003
$p_{TG 2, N=10}$	2	3	$p_{TG 3, N=24}$	2	3
1	0,005	0,002	1	< 0,001	< 0,001
2	–	< 0,001	2	–	< 0,001
Legende: vgl. Tabelle 8a					

Tab. 8b: Signifikante Korrelationen des Korrelationskoeffizienten r nach Pearson zwischen den Messzeitpunkten 1 – 3 der Experimentalgruppen für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests. **Fettdruck:** signifikante Effekte für $p < 0,05$

Bezüglich der Teststärke bietet eine Varianzanalyse mit Messwiederholung (RM-ANOVA) gegenüber mehreren t-Tests für abhängige Stichproben keinen Vorteil, weil beide Berechnungsverfahren an denselben Messdaten durchgeführt werden müssen, so dass die Stichprobengröße beim Vergleich über alle Messzeitpunkte von einem Wechsel der Auswertungsverfahren unberührt bleibt. Obwohl eine RM-ANOVA durch den einmaligen Vergleich mehrerer Mittelwerte vom Problem der α -Fehler-Kumulierung befreit, darf bei einer Varianzanalyse mit Messwiederholung die Abhängigkeit der Messdaten von den verschiedenen Messzeitpunkten nicht unterschlagen werden. Denn eine RM-ANOVA ist gegenüber einer ANOVA zusätzlich an die Voraussetzung gebunden, dass alle Korrelationen zwischen den Stufen des messwiederholenden Faktors homogen sein müssen. Wenn die Korrelationen nicht homogen sind, dann kann die Signifikanzprüfung einer einfaktoriellen, messwiederholenden Varianzanalyse fälschlicherweise zu einem signifikanten Ergebnis führen (Rasch & Frieze, 2006). Eine progressive Entscheidung zugunsten der Alternativhypothese kann nach den Ergebnissen der Tabellen 8a und 8b nicht ausgeschlossen werden. Mit Ausnahme der Testgruppe Reorganisation (TG 2) geben die Ergebnisse zu erkennen, dass die Korrelationen zwischen den Messzeitpunkten 2 und 3 sinken und zwischen den Messzeitpunkten 1 und 3 ein Minimum erreichen. Die Ergebnisse boten den Anlass, die Unterschiede zwischen den Messzeitpunkten für die Dimension Naturwissenschaftliche Arbeitsweisen des Fach- & NAW-Tests nicht durch Varianzanalysen für abhängige Stichproben, sondern nur durch t-Tests für abhängige Stichproben zu dokumentieren. Es muss jedoch angemerkt werden, dass die Gesamtwahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl sie in Wahrheit gilt, bei paarweisen Vergleichen von drei Mittelwerten mit den drei Kombinationen für die Messzeitpunkte 1/ 2, 2/ 3 und 1/ 3 und einem pro Test festgelegtem α -Fehlerniveau von 5 % auf 14 % erhöht werden kann.

4.4 Ergebnisse der Beobachtung

4.4.1 Prozessdiagramme und Prozessanalyse des Experimentierens

Zur Analyse der Experimentierprozesse werden zunächst die Beobachtungsprotokolle von den 41 Probandenpaaren der Teilstichprobe mit Hilfe des entwickelten Codiersystems (vgl. Kap. 4.2.3) in Prozessdiagramme (Synonym: Prozessgrafik) transformiert. In einem Folgeschritt werden die Prozessgrafiken in Niveaustufendiagramme überführt (vgl. Kap. 4.4.3). Die Niveaudiagramme dienen

der Bewertung und der Einweisung der Experimentiertätigkeiten in Niveaustufen. Da dieses Vorgehen der Datenaufbereitung eine detaillierte Analyse der Prozessgrafiken voraussetzt, soll in diesem Kapitel die Auswertung der erstellten Prozessdiagramme exemplarisch aufgezeigt werden. Um aus der Gesamtheit der Teilstichprobe die Variationsbreite bzw. einen möglichst großen Abstand zwischen den beobachtbaren, empirisch gefundenen Experimentierprozessen für den Umgang mit Variablen zu dokumentieren, werden mit einer gezielten Auswahl zwei Experimentierabläufe von Probandenpaaren kontrastiert, die vor der Bearbeitung der offenen Experimentieraufgabe eine bzw. keine vorbereitende Instruktion für den Umgang mit Variablen erhalten haben. Zum Vergleich werden ein Prozessdiagramm aus der Testgruppe 1 (Reproduktion) als ein Beispiel für einen strukturierten Experimentierverlauf (vgl. Abbildungen 8 a und b) und ein Prozessdiagramm aus der Kontrollgruppe als ein Beispiel für einen nicht strukturierten Experimentierverlauf (vgl. Abbildungen 9 a und b) gegenübergestellt. Den beiden exemplarischen Prozessanalysen wird ein Abschnitt vorangestellt, indem das prinzipielle Vorgehen zur Auswertung der Prozessgrafiken thematisiert wird. Zum Abschluss des Kapitels werden ergänzende und kritische Anmerkung zur Erstellung und Analyse der Prozessgrafiken gemacht.

Prinzipielles Vorgehen zur Analyse der Prozessdiagramme. Der Aufbau der Prozessanalysen orientiert sich an der Zeitachse der Prozessdiagramme. Ihr Informationsgehalt liegt in der Abbildung der probandenpaarspezifischen Sequenzierung der Experimentierzeit, die zur Bearbeitung der offenen Experimentieraufgabe 150 Minuten nicht überschreiten darf. Die Experimentiersequenzen sind an den numerischen Codierungen in den Prozessdiagrammen zu erkennen (vgl. Kap. 4.2.3). Sie verschlüsseln die Kategorien von fünf beobachtbaren Experimentiertätigkeiten. Ziffernfolgen, die mit einer 1, 2 oder 3 beginnen, zeigen beobachtbare Experimentiertätigkeiten der ersten Experimentierphase an. Hierzu gehören das Beschaffen von Informationen (Ziffer 1), die Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset (Ziffer 2) und die Planung von Profilkonstruktionen (Ziffer 3). Ziffernfolgen, die mit einer 4 oder 5 beginnen, zeigen beobachtbare Experimentiertätigkeiten der zweiten Experimentierphase an. Hierzu gehören die Fertigung eigener Profile (Ziffer 4) und die Änderung an einer bestehenden Profilkonstruktion (Ziffer 5). Mit dieser Festlegung der Codiervorschrift ist es möglich, dass aus den Prozessdiagrammen sowohl die Häufigkeit als auch der zeitliche Umfang von Experimentiersequenzen wie der Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset oder der Konstruktion eigener Profile entnommen werden kann. Außerdem ermöglicht das entwickelte Codiervorgehen, dass Übergänge bzw. Wechsel zwischen den Experimentierphasen 1 und 2 sehr einfach identifiziert werden können. Um die Arbeit mit den Flügelsets oder die Konstruktion eigener Profile korrekt auswerten zu können, muss allerdings bekannt

sein, dass der mehrmalige Gebrauch verschiedener Sets oder die Konstruktion mehrerer Profile durch Wiederholung von numerischen Folgen dargestellt werden, die mit derselben Ziffer beginnen. Darüber hinaus werden im Prozessdiagramm die Profile, die von den Probandenpaaren selbst konstruiert wurden nummeriert und mit dem Zusatz P (Profil) gekennzeichnet (z. B. Abbildung 8b, P₁). Die Markierung soll gewährleisten, dass potentielle Änderungen an bestehenden Profilen den korrespondierenden Eigenkonstruktionen zugeordnet werden können. Mit diesem Verfahren können in der zweiten Experimentierphase Messfolgen identifiziert werden.

Die beobachtbaren und numerisch codierten Experimentierabläufe sind im Allgemeinen mit den verbalen Argumentationen der Probandenpaare vernetzt. Diese „Tonspur“ wird in den Prozessdiagrammen symbolisch mit einem Pfeil codiert. Jeder Pfeil wird durch einen alphabetischen Code ergänzt, der die Hauptkategorien der verbalen Argumentation anzeigt. Wie dem Kapitel 4.2.3 zu entnehmen ist, steht ein „V“ für die Kategorie Vergleich und Bewertung von Messfolgen, ein „S“ für die Kategorie Idee und Strategie zur Profilkonstruktion und ein „E“ für die Kategorie Erklärung. Für die Auswertung der Prozessdiagramme muss bekannt sein, dass ein Pfeil nur den Zeitpunkt und nicht die zeitliche Länge einer verbalen Argumentation markiert. Ausnahmen liegen vor, wenn von den Probandenpaaren in kurzer Folge mehrere verbale Argumentationen gemacht werden. In diesem Fall wird auf eine separate Kennzeichnung mit einem eigenen Pfeil zugunsten einer zusammenfassenden Darstellung von mehreren codierten Kategorien oberhalb eines Pfeils verzichtet (z. B. Abbildung 8b, Nr. 11). Damit der Aufbau einer Prozessanalyse übersichtlicher gestaltet werden kann, wird jeder Pfeil über den alphabetischen Code hinaus mit einer fortlaufenden Nummer unterlegt.

Das bis jetzt vorgestellte Auswertungsverfahren ist geeignet, um aus den Prozessgrafiken den Verlauf von beobachtbaren Experimentiersequenzen, die von verbalen Argumentationen begleitet werden, zu sichten. Eine tiefergehende Analyse der Experimentierprozesse erfordert jedoch, dass die numerisch codierten Informationen (beobachtbare Experimentiertätigkeiten) mit den symbolisch codierten Informationen (verbale Argumentationen) gekreuzt werden. Hierin liegt ein Schwerpunkt der folgenden, exemplarischen Prozessanalysen. Ein weiterer Schwerpunkt der exemplarischen Prozessauswertungen besteht darin, dass Aussagen über die Qualität der Experimentierprozesse gemacht werden. Aus diesem Grund darf bei den numerischen Codierungen nicht nur die erste Ziffer und bei den verbalen Argumentationen nicht nur die Hauptkategorie berücksichtigt werden. Mit einem Verweis auf das Kapitel 4.2.3 sei daran erinnert, dass die Indizes der symbolischen Codierungen (z. B. V₃) die Qualität der verbalen Argumentationen kennzeichnen.

Exemplarische Prozessanalyse eines Probandenpaares der Testgruppe 1.

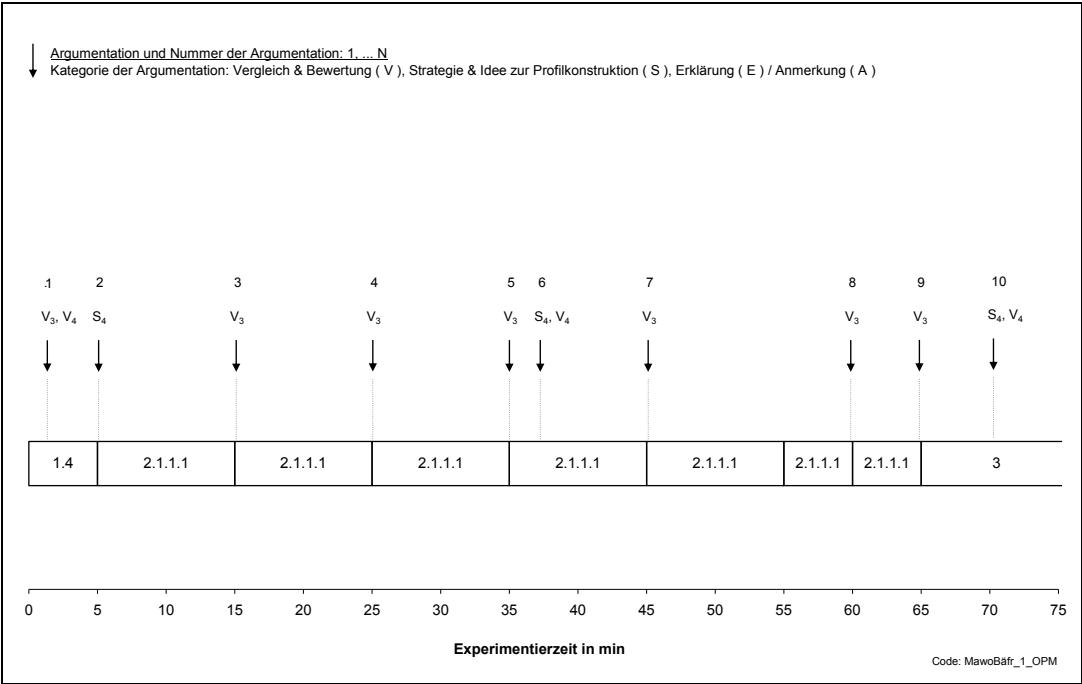


Abb. 8a: Prozessgrafik eines Probandenpaares der Testgruppe 1 (Reproduktion)

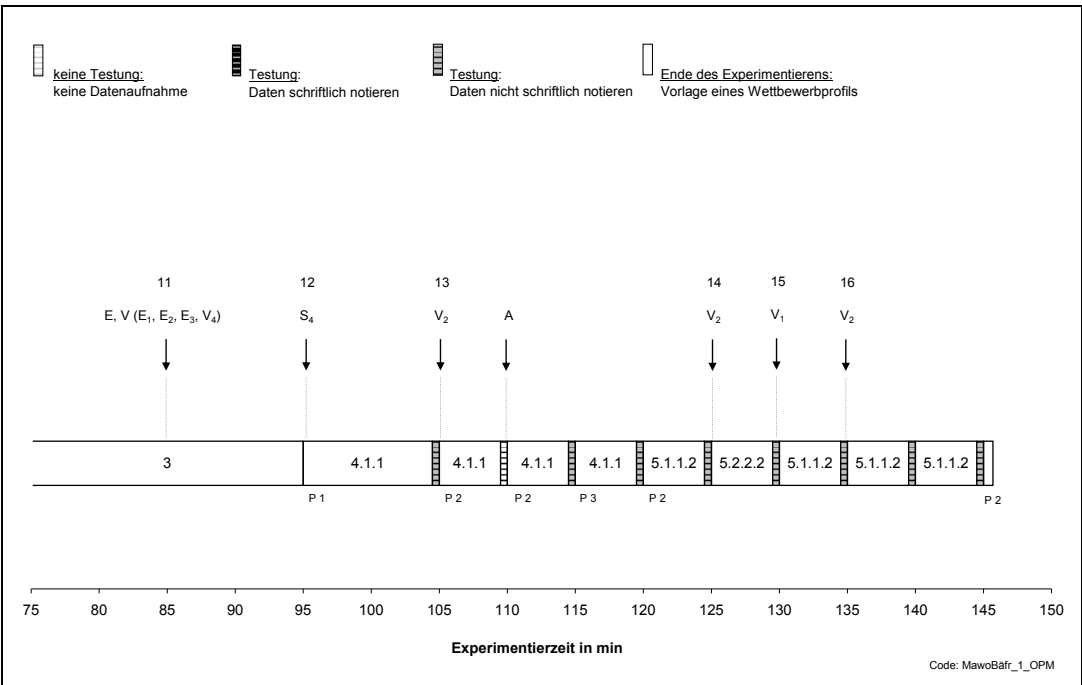


Abb. 8b: Fortsetzung der Prozessgrafik eines Probandenpaares der Testgruppe 1 (Reproduktion)

Zum ersten Prozessdiagramm gehören die Abbildungen 8a und 8b, die den Experimentierverlauf eines Probandenpaares wiedergeben, das vor der Durchführung der offenen Experimentieraufgabe eine vorbereitende Instruktion zum Umgang mit Variablen erhalten hat. Das Diagramm 8b ist eine Weiterführung der Abbildung 8a. Beide Diagramme sind durch die Zeitachse miteinander verbunden. Um die Auswertung der Prozessgrafik nicht nur anhand der Zeitmarken verfolgen zu können, wird in der experimentierphasenspezifischen Prozessanalyse punktuell auf die entsprechenden Codierungen verwiesen.

Experimentierphase 1 der Testgruppe 1 (0 min – 95 min). Der Abbildung 8a ist zu entnehmen, dass das Probandenpaar zum Einstieg in die offene Experimentieraufgabe auf das schriftliche Material der vorbereitenden Instruktion zurückgreift (Code 1.4). Die äußere Handlung der fünf Minuten dauernden Sequenz gleicht einer Wiederholung von zentralen Ergebnissen. Inhaltlich befasst sich das Probandenpaar mit der Auswertung der präsentierten Messfolgen, indem Vergleiche und Bewertungen einer bzw. mehrerer Messfolgen durchgeführt und Schlussfolgerungen auf die Flügeleigenschaft gemacht werden (Nr. 1; Codes V₃, V₄). Am Ende der Informationsbeschaffungsphase wird erstmals eine Strategie und Idee zur Profilkonstruktion verbalisiert (Nr. 2; Code S₄). Das Probandenpaar verfolgt das Ziel, mit der eigenen Profilkonstruktion die Bedingungen zur Erzeugung von maximalen Spannungen mehrerer Variablen zu vereinen („Vereinigung von Maximalwerten“). Weiterhin ist aus der Codierung zu lesen, dass eine Spezifizierung der Strategie vorliegt: Über die Vereinigung der Bedingungen von maximalen Spannungswerten hinaus, sollen Variablen mit einer großen Effektstärke bestimmt und in eine Rangfolge gebracht werden. Dem Prozessdiagramm ist zu entnehmen, dass diese Strategie im Verlauf der Experimentiertätigkeiten mit der Aufnahme von Messfolgen unter Einsatz des aufgabenspezifischen Experimentiermaterials Flügelsets weiter präzisiert wird (Nr. 6, 10, 12; Code S₄). Eine Änderung der Strategie, die durch einen Wechsel der Codierung erkannt werden würde, z. B. von S₄ nach S₂, liegt nicht vor.

Um die Strategie bei den eigenen Profilkonstruktionen umsetzen zu können, müssen vorab nicht konfundierende Messfolgen aufgenommen, schriftlich fixiert und ausgewertet werden. Diese Voraussetzungen werden von dem Probandenpaar erfüllt, was an der systematischen Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelsets (Code 2.1.1.1) und den nicht konfundierenden Vergleichen und Bewertungen der empirisch gewonnen Messdaten zu erkennen ist (Nr. 3 – 5, 7 – 9; Code V₃ und Nr. 6, 10, 11; Code V₄).

Die sich an die Aufnahme und Auswertung der Messfolgen anschließende Phase der Profilplanung wird von einer Diskussion zwischen den Probanden dominiert. Die Diskussion wird vor allem von Plausibilitätserklärungen (Code E₁), Vermutungen über den Einfluss weiterer, bisher nicht getesteter Variablen (Code E₂) und

fachlichen Erklärungen (Code E₃) beherrscht. Darüber hinaus geht es in der Diskussion auch um die Planung einer geeigneten Profilgrundform, mit der die Bedingungen zur Erzeugung von maximalen Spannungen vereinigt werden sollen (s.o.). Obwohl die Erklärungen mit Qualitätsschwankungen behaftet sind (vgl. E₁, E₃), ist die Lösung der Planungsphase überzeugend: Die besten Bedingungen sollen an drei verschiedenen Profilgrundformen erprobt werden. Anschließend soll die Profilgrundform mit den besten Eigenschaften durch weitere Änderungen optimiert werden.

Experimentierphase 2 der Testgruppe 1 (95 min – 145 min). Zum Einstieg in die Experimentierphase 2 werden die drei geplanten Profilgrundformen konstruiert und getestet (P₁ – P₃; Code 4.1.1). Die Codierungen zeigen an, dass es sich bei allen Profilen um messdatenbasierte Konstruktionen handelt, die durch eine Kombination von mehreren Parametern gewonnen werden. Das Profil P₂ wird zweimal konstruiert, weil es beim ersten Fertigungsversuch zerbricht. Diese Information kann der mit A versehenen Anmerkung entnommen werden. Nach Abschluss der Profilkonstruktionen werden an dem Profil P₂ fünf Änderungen vorgenommen. Bis auf eine Ausnahme handelt es sich um einmalige, zueinander nicht konfundierende Änderungen von Variablen, die ohne Kontrolle der Ausprägung variiert werden (Code 5.1.1.2). Eine der Variablen wird an dem Profil P₂ wiederholt verändert. Allerdings ist die Ausführung konfundierend (Code 5.2.2.2), so dass die Aufnahme einer Messfolge verhindert wird. In Analogie zur ersten Experimentierphase werden auch in der Experimentierphase 2 die erhobenen Messdaten verglichen und bewertet. Jedoch ist mit dem Phasenübergang ein Wechsel der Kategorien von V₃ bzw. V₄ (Phase 1) nach V₁ bzw. V₂ (Phase 2) festzustellen. Die Codierung V₂, die häufigste Vergleichskategorie in der Experimentierphase 2, besagt, dass es sich um Vergleiche und Bewertungen von Messwertpaaren mit Schlussfolgerung auf die bessere Flügeleigenschaft handelt (Nr. 13, 14, 16). Dieses Vorgehen wird bis zur Vorlage der optimierten Profilkonstruktion fortgesetzt. Es fällt auf, dass die Messergebnisse in dem aus der Experimentierphase 1 bereits vorhandenen Messprotokoll nicht mit aufgenommen werden (vgl. Abbildung 8b, Testung: Messdaten nicht schriftlich notieren).

Exemplarische Prozessanalyse eines Probandenpaares der Kontrollgruppe. Zum Prozessdiagramm der Kontrollgruppe gehören die Abbildungen 9a und 9b. Die Abbildungen dokumentieren den Experimentierverlauf eines Probandenpaares, dass vor der Bearbeitung der offenen Experimentieraufgabe keine vorbereitende Instruktion zum Umgang mit Variablen erhalten hat.

Eine erste Sichtung der Abbildungen 9a und 9b ergibt, dass die Experimentierphasen 1 und 2 nacheinander durchlaufen werden. Das Probandenpaar gestaltet die Phase 1 ausschließlich durch die Arbeit mit dem aufgabenspezifischen Experimentier-

4 HAUPTSTUDIE

material Flügelsets. Phasen der Informationsbeschaffung oder der Profilplanung kommen nicht vor.

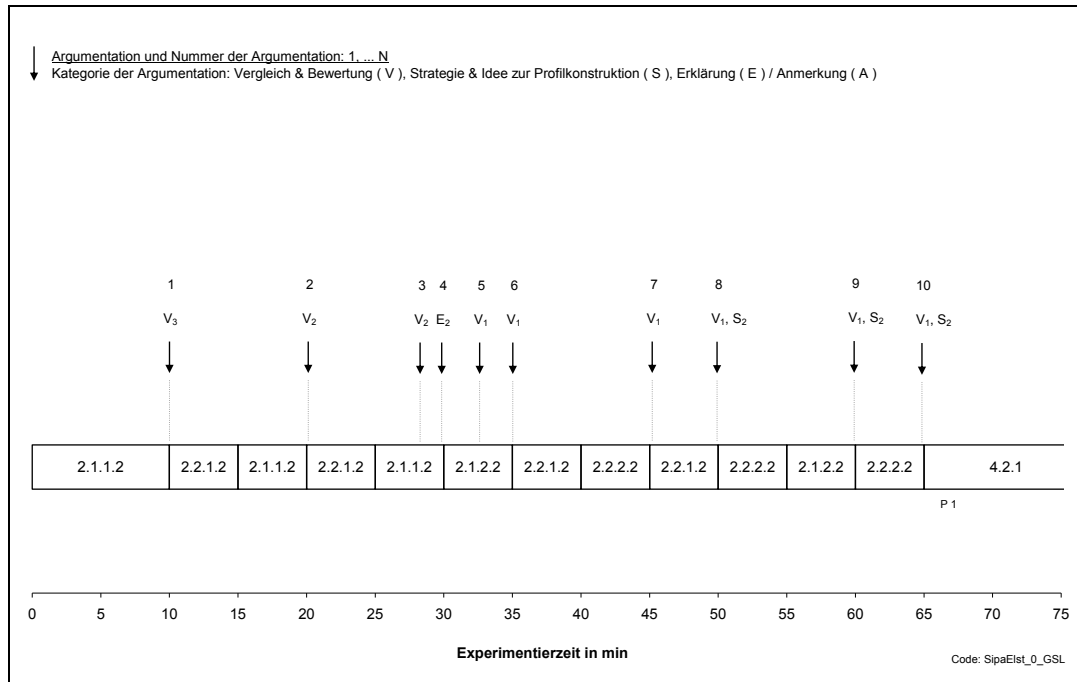


Abb. 9a: Prozessgrafik eines Probandenpaares der Kontrollgruppe

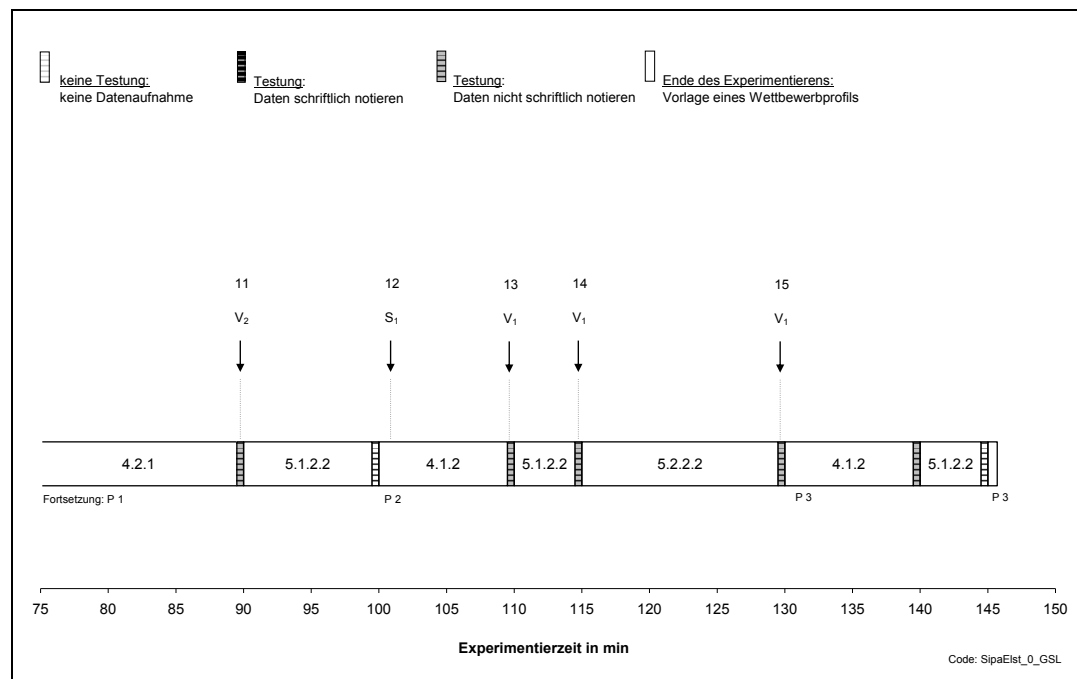


Abb. 9b: Fortsetzung der Prozessgrafik eines Probandenpaares der Kontrollgruppe

Bemerkenswert ist, dass das Probandenpaar mit 12 Flügelsets experimentiert, obwohl nur Messfolgen mit 8 verschiedenen Variablen zur Verfügung gestellt werden. Messwiederholungen können ausgeschlossen werden, da andernfalls die numerischen Codierungen zusätzlich mit einem W (Wiederholung) markiert worden wären. Das Probandenpaar muss über den Gebrauch der angebotenen Messfolgen hinaus, einen oder mehrere Profile aus den Flügelsets separiert und neue Flügelsetkombinationen zusammengestellt haben. Nach Abschluss der Arbeit mit dem aufgabenspezifischen Experimentiermaterial werden in der zweiten Phase drei verschiedene Profile (Code P_1 , P_2 , P_3) konstruiert, die nach ihrer Fertigstellung durch ein bis zwei bauliche Änderungen noch weiter optimiert werden. Weiterhin ist den Prozessdiagrammen zu entnehmen, dass die Wirkung der baulichen Änderungen von dem Probandenpaar nicht immer überprüft wird (vgl. Abbildung 9b, keine Testung).

Auf der Ebene der verbalen Argumentationen ist vor allem die Kategorie Vergleich und Bewertung zu finden. Zudem können direkt vor dem Übergang zur Experimentierphase 2 mehrfach verbale Argumentationen der Kategorie Strategie und Idee zur Profilkonstruktion registriert werden. Aufgrund des späten Zeitpunkts, 50 Minuten nach Beginn der Experimentiertätigkeit, liegt die Vermutung nahe, dass der Experimentierprozess ohne ein strategisches Ziel begonnen wurde. Darüber hinaus ist bei diesem Probandenpaar ein Mangel an Erklärungen festzustellen.

Nach dieser ersten Sichtung der Prozessgrafiken folgt nun eine tiefergehende, experimentierphasenspezifische Prozessanalyse, in der punktuell auf die entsprechenden Codierungen verwiesen wird.

Experimentierphase 1 der Kontrollgruppe (0 min – 65 min). Auf der Ebene der beobachtbaren Experimentiertätigkeiten können aus den Codierungen drei Bereiche identifiziert werden. Im ersten Bereich (0 min – 30 min) werden mit den Flügelsets nur nicht konfundierende Messungen bzw. Messfolgen aufgenommen. Im zweiten Bereich (30 min – 50 min) werden konfundierende und nicht konfundierende Experimentierweisen im Wechsel ausgeführt, wohingegen das Probandenpaar im dritten Bereich (50 min – 65 min) nur noch konfundierend arbeitet. Der Übergang von der nicht konfundierenden zur konfundierenden Arbeitsweise geht mit der verbalen Argumentation der Kategorie E_2 einher (Nr. 4). Sie steht für eine Vermutung über den Einfluss einer bisher nicht getesteten Variablen. Vor dieser Schlüsselargumentation nimmt die Gruppe nur einmal eine nicht konfundierende Messfolge (Code 2.1.1.2), zweimal eine Messfolge mit nur zwei Messwerten (V_2 mit Code: 2.1.1.2) und zweimal Einzelmessungen auf (Code 2.2.1.2). Nach der Schlüsselargumentation E_2 folgen mit Ausnahme der beiden nicht konfundierenden Experimentiersequenzen im zweiten Bereich, die Aufnahmen von konfundierenden Messfolgen (Code 2.1.2.2) und konfundierenden Einzelmessungen (Code 2.2.2.2).

Aber nicht nur auf der Ebene der beobachtbaren Experimentiertätigkeiten, sondern auch auf der Ebene der verbalen Argumentationen geht mit der Vermutung E₂ ein Qualitätsverlust des Experimentierens einher. Er ist an dem Wechsel der Kategorie Vergleich und Bewertung von Messfolgen durch den Übergang der Argumentationen von V₂ (Nr. 2, 3) nach V₁ (Nr. 5 – 10) zu bemerken. Die Codierung V₁ besagt, dass es sich um Vergleiche konfundierender Messergebnisse handelt. Bemerkenswert ist, dass diesem Wechsel schon ein Qualitätsabfall vorausgeht, denn im Bereich der nicht konfundierenden Experimentiersequenzen (0 min – 30 min) findet im zeitlichen Abstand vor dem Schlüsselargument E₂ ein Übergang der Argumentationen von V₃ (Nr. 1) zu V₂ (Nr. 2, 3) statt. Zu allen Messungen mit dem aufgabenspezifischen Experimentiermaterial ist anzumerken, dass das Probandenpaar keines der zahlreichen Ergebnisse schriftlich notiert. Die beabsichtigte Strategie der Profilkonstruktion, das beste Profil der Flügelsets zu kopieren (Nr. 8 – 10; Code S₂), wird dadurch erschwert.

Experimentierphase 2 der Kontrollgruppe (65 min – 145 min). Der Einstieg in die Phase 2 wird durch die Umsetzung der Strategie S₂ mit der Profilkonstruktion P₁ bestimmt. Der Eigenbau stimmt exakt mit einem Profil der Flügelsets überein (Code 4.2.1). In Abgrenzung dazu handelt es sich bei den anderen beiden Profilen um Konstruktionen, die willkürlichen Parameterkombinationen entsprungen sind und nicht auf einer messdatenbasierten Arbeitsweise fußen (Code 4.1.2; P₂, P₃). Im Gegensatz zur Profilkonstruktion P₁ wird dieser Strategiewechsel durch die Codierung S₁ (Nr. 12) angezeigt. Die baulichen Änderungen an den Profilen sind alle konfundierend (Code 5.1.2.2) und in einem Fall entsteht aus den Änderungen an der Profilkonstruktion P₂ sogar eine konfundierende Messfolge (Code 5.2.2.2). Auf der Ebene der verbalen Argumentationen überwiegen die Vergleiche und Bewertungen von Messfolgen mit der Kategorie V₁ (Nr. 13 – 15). Der einmalige paarweise Vergleich zweier Messergebnisse (Nr. 11; Code V₂) ist als sequenziell begrenzte Qualitätssteigerung zu bewerten.

Ergänzende und kritische Anmerkungen zur Erstellung und Analyse der Prozessgrafiken. Die folgenden drei Ergänzungen und kritischen Anmerkungen beruhen auf Erkenntnissen, die aus den Prozessanalysen der Teilstichprobe (41 Probandenpaare) erhalten wurden. Aus diesem Grund dienen mögliche Verweise auf die exemplarischen Prozessgrafiken der Abbildungen 8a und b sowie den Abbildungen 9a und b nur der Illustration.

Zur Auswertung der Prozessgrafiken muss erstens angemerkt werden, dass bei den meisten Probandenpaaren der Teilstichprobe die Anzahl der verbalen Argumentationen mit Beginn der zweiten Experimentierphase stagniert. Im Extremfall, der auch für die Abbildungen 8b und 9b zutreffend ist, tauschen die Probandenpaare über längere Zeiträume überhaupt keine verbalen Argumentationen mehr aus (z. B. Abbildung 8b, 105 min – 125 min). Weiterhin gehört zu dem experimental-

gruppenübergreifenden Effekt, dass unter den registrierten verbalen Argumentationen in der Experimentierphase 2 vor allem die Kategorie Vergleich und Bewertung von Messfolgen zu finden ist. Die Kategorien Idee und Strategie zur Profilkonstruktion sowie Erklärung sind hauptsächlich in der Experimentierphase 1 vertreten. Um einen Datenverlust als Ursache des empirischen Effekts auszuschließen, kann als Ergebnis einer Methodentriangulation (videografische Beobachtung, Beobachtungsprotokolle, Probandenbefragung) berichtet werden, dass die Experimentierenden zu einer Auslassung von Vergleichen und Bewertungen neigen, wenn sie sich über eine Strategie für die eigene Profilkonstruktion verständigt haben. In diesem Fall werden die Vergleiche von Messergebnissen nicht mehr wie in der Experimentierphase 1 explizit verbalisiert, sondern simultan mit dem Ablesen vom Messinstrument vollzogen. Das hat zur Konsequenz, dass z. B. für die Argumentationen 13 oder 16 der Kategorie V_2 in der Abbildung 8b die Lesart Bereich der paarweisen Vergleiche von zwei Messergebnissen zutreffender ist als die Lesart, dass das Probandenpaar 105 min bzw. 125 min nach Beginn der Experimentierzeit zwei Messergebnisse paarweise miteinander verglichen und bewertet hat. Angemerkt sei, dass trotz der unregelmäßigeren Verteilung der verbalen Argumentationen in der Experimentierphase 2 die Anzahl und Qualität der Vergleiche den Kriterien zur Bewertung und Einweisung der Experimentierprozesse in die Niveaustufen des Experimentierens genügen (vgl. Kap. 4.4.2).

Zweitens muss zur Auswertung der Prozessgrafiken angemerkt werden, dass sich unter den 41 Probandenpaaren der Teilstichprobe 8 Probandenpaare befinden, deren Experimentierprozesse nicht nur von trainierten Beobachtern protokolliert, sondern zusätzlich auch videografiert wurden. Die Herkunft der 8 Probandenpaare ist über die Experimentalgruppen gleichmäßig verteilt. Das Ziel der Maßnahme ist eine punktuelle Kontrolle der Messdaten (zeitliche Anordnung, Eindeutigkeit der Kategoriezuordnungen und Vollständigkeit der Beobachtungen) im Längsschnitt der Studie. Als Ergebnis dieser methodischen Variante der Beobachterübereinstimmung kann berichtet werden, dass die Mitschriften der Datenaufnehmer durch die videografischen Aufnahmen gestützt werden. Aus diesem Grund wird auf eine explizite Trennung und Nennung von Prozessdiagrammen verzichtet, deren Datenaufnahmeprozess auf zwei unabhängigen Erhebungsmethoden basiert.

Mit einem Verweis auf das Kapitel 4.4.2 soll abschließend angemerkt werden, dass die Kategorie Idee und Strategie für die eigene Profilkonstruktion nicht in den Kriterienkatalog zur Einweisung der Experimentierprozesse in die Niveaustufen des Experimentierens mit aufgenommen wird. Die Entscheidung basiert auf den 41 Prozessanalysen der Teilstichprobe, die ergaben, dass diese Kategorie der verbalen Argumentation vor allem in der Experimentierphase 1 auftritt. Das Ergebnis ist nicht mit den Kriterien für die Bewertung der Experimentierprozesse zu vereinen, weil für

beide Experimentierphasen dieselben Bewertungsgrundlagen gelten sollen (vgl. Kap. 4.2.3).

4.4.2 Kriterien zur Bewertung und Einweisung der Experimentierprozesse in Niveaustufen

Die Kriterien zur Bewertung und Einweisung der Experimentierprozesse in Niveaustufen werden aus den beobachtbaren Experimentiertätigkeiten und den verbalen Argumentationen der Probandenpaare abgeleitet. Für die Aufstellung der Niveaustufen ist ausschlaggebend, dass jede Niveaustufe aus einer Kombination von beobachtbaren Experimentiertätigkeiten und verbalen Argumentationen besteht. Weiterhin ist zu beachten, dass die gewählten Kategorien der beobachtbaren Experimentiertätigkeiten und verbalen Argumentationen auch Bestandteil der vorbereitenden Instruktion waren. Aus diesem Grund werden z. B. Strategieausprägungen für die eigene Profilkonstruktion wie die Vereinigung von Bedingungen zur Erzeugung von maximalen Spannungswerten mehrerer Variablen oder die exakte Kopie eines Profils der Flügelsets nicht mit in den Kriterienkatalog aufgenommen. Ein weiterer wichtiger Aspekt zur Festlegung der Niveaustufen ist, dass für die beiden Experimentierphasen dieselben Bewertungsgrundlagen gelten. In der Summe führen die Kriterien zu einer Einteilung der Experimentierprozesse in drei Niveaustufen. Die unterste, unsystematische Niveaustufe wird als Stufe 1, die mittlere, semi-systematische Niveaustufe wird als Stufe 2 und die oberste, systematische Niveaustufe wird als Stufe 3 bezeichnet. Eine zusammenfassende Übersicht über die Einteilung der Kriterien und Kategorien der Niveaustufen gibt die Tabelle 9. Der Tabelle ist zu entnehmen, dass die Abgrenzung der Niveaustufen 1 und 2 unter anderem durch eine konfundierende und nicht messdatenbasierte bzw. durch eine nicht-konfundierende und messdatenbasierte Arbeitsweise erfolgt. Der Tabelle ist weiterhin zu entnehmen, dass die nicht-konfundierende und messdatenbasierte Experimentierweise nicht nur für die Niveaustufe 2, sondern auch für die Niveaustufe 3 kennzeichnend ist. Die Abgrenzung zwischen den beiden Niveaustufen liegt in der Ausprägung der experimentellen Tätigkeiten. Für den Fall, dass die Experimentierenden nur Ansätze von Messfolgen (einfache Messfolgen mit nur zwei Messwerten) aufnehmen, vergleichen und bewerten wird eine Eingruppierung in die Niveaustufe 2, andernfalls in die Niveaustufe 3 vorgenommen.

Mit einem Ausblick auf die Vergabe der Niveaustufen (vgl. Kap. 4.4.3 f.) sei angemerkt, dass die Probandenpaare beobachtbare Experimentiertätigkeiten ausführen können, die niveauvoller sind als die verbalen Argumentationen, die die Experimentiertätigkeiten begleiten. Als ein Beispiel sei die Aufnahme von nicht-konfundierenden Messfolgen (Kriterium der Niveaustufe 3) mit paarweisen Vergleichen der Messergebnisse (Kriterium der Niveaustufe 2) genannt. Wenn wie in

dem beschriebenen Beispiel zwischen den beobachtbaren Experimentiertätigkeiten und den verbalen Argumentationen Niveaustufenunterschiede vorliegen sollten, dann richtet sich die Vergabe der Gesamt-Niveaustufe nach den Kriterien der tieferen Niveaustufe. In diesem Fall wird auf eine separate Vergabe von Niveaustufen für die verbalen Argumentationen und beobachtbaren Experimentiertätigkeiten der Probandenpaare verzichtet, weil die verbalen Argumentationen nicht isoliert, sondern stets als Begleiter der experimentellen Handlungen auftreten. Dieses Verfahren zur Einweisung der Experimentierprozesse in Niveaustufen ist dem Anspruch geschuldet, eine möglichst umfassende Bewertung der Experimentierprozesse zum Umgang mit Variablen vorzunehmen.

Niveaustufe 1 (unsystematisch)	
Tätigkeiten des Experimentierens	Messfolge oder Ansatz einer <i>konfundierenden</i> Messfolge aufnehmen und/ oder <i>nicht-konfundierende</i> bzw. <i>konfundierende</i> Einzelmessung aufnehmen; Profil konstruieren: <i>nicht messdatenbasiert</i>
Vergleich und Bewertung	(V ₁) Bewertung ohne Vergleich bzw. Vergleich und Bewertung <i>konfundierender</i> Messergebnisse
Niveaustufe 2 (semi-systematisch)	
Tätigkeiten des Experimentierens	Ansatz einer <i>nicht-konfundierenden</i> Messfolge aufnehmen; Profil konstruieren: <i>messdatenbasiert</i>
Vergleich und Bewertung	(V ₂) Vergleich eines Messwertes mit einem anderen und Schlussfolgerung auf die bessere Flügeleigenschaft
Niveaustufe 3 (systematisch)	
Tätigkeiten des Experimentierens	<i>nicht-konfundierende</i> Messfolge aufnehmen; Profil konstruieren: <i>messdatenbasiert</i>
Vergleich und Bewertung	(V ₃) Vergleich einer Messfolge mit Bewertung, (V ₄) Vergleich mehrerer Messfolgen mit Bewertung

Tab. 9: Kriterien und Kategorien zur Bewertung und Einweisung der Experimentierprozesse in Niveaustufen

4.4.3 Experimentiermuster und Experimentiertypen

Die Identifizierung von charakteristischen Experimentiermustern und Experimentiertypen geschieht durch eine Auswertung der Niveaustufengrafiken von allen Probandenpaaren der Teilstichprobe. Für den Umgang mit Variablen werden die Niveaustufengrafiken des Experimentierens durch eine Bewertung der Experimentierprozesse generiert, die mit Hilfe der Prozessgrafiken beschrieben werden (vgl. Kap. 4.4.1). Zur Bewertung der Experimentierprozesse werden zwei Kriterien, die Niveaustufe des Experimentierens und die Niveaustufenstabilität des Experimentierens, herangezogen. Das erste Kriterium wird als notwendig erachtet, damit die Qualität des Experimentierens in einem hierarchischen Ordnungsprinzip abgebildet werden kann; das zweite Kriterium, um feststellen zu können, ob die Experimentierprozesse der Probandenpaare mit Qualitätsunterschieden behaftet sind.

Die Überführung der Prozess- in die Niveaustufengrafiken folgt einem zweistufigen Verfahren: Zuerst werden die Experimentiersequenzen der Probandenpaare auf Grundlage der im Kapitel 4.4.2 aufgestellten Kriterien zur Einweisung der Experimentierprozesse in die Niveaustufen des Experimentierens bewertet. Anschließend wird unter Anwendung des Kriteriums „Niveaustabilität“ geprüft, ob die Probandenpaare in den beiden Experimentierphasen niveaustabil oder niveauinstabil experimentiert haben. Wenn im Verlauf der Experimentierphase 1 oder der Experimentierphase 2 keine bzw. mit einem Toleranzbereich behaftete Niveaustufenwechsel auftreten, dann wird die korrespondierende Phase mit dem Merkmal „stabil“, andernfalls mit dem Merkmal „instabil“ bewertet. Die Toleranzbereiche können den Experimentiertypenbeschreibungen entnommen werden (s. u.).

Zur Identifizierung der Experimentiertypen sind aus den Kriterien Niveaustufe (mit den Ausprägungen unsystematisch, semi-systematisch oder systematisch) und Niveaustabilität (mit den Ausprägungen niveaustabil oder niveauinstabil) zahlreiche, aber nur endlich viele Typenkombinationen möglich. Wie ein Ausblick auf die Ergebnisse zeigt, konnten empirisch nur vier typische Experimentiermuster gefunden werden (vgl. Abbildung 10). Diese werden mit A – D bezeichnet und folgendermaßen charakterisiert: Experimentiertyp A (hohes Experimentierniveau, niveaustabil), Experimentiertyp D (niedriges Experimentierniveau, niveauinstabil). Die beiden anderen Typen können gemäß ihrer alphabetischen Bezeichnung in die hierarchische Ordnung eingefügt werden. Die Bezeichnung „typisches Experimentiermuster“ wurde vergeben, wenn von den 41 Probandenpaaren der Teilstichprobe, mindestens vier Probandenpaare (10 % der Probandenpaare) demselben Muster folgten.

Experimentierphasenspezifische Charakterisierung der Experimentiertypen mit den Kriterien Niveaustufe und Niveaustufenstabilität			
Typ	Phase	Niveaustufe	Stabilität
A	1	3	niveaustabil
	2	2	niveaustabil
B	1	3	niveaustabil
	2	1	niveaustabil
C	1	2	niveaustabil
	2	1	niveaustabil
D	1	2	niveaustabil
	2	1	niveaustabil oder -instabil
D	1	1	niveaustabil oder -instabil
	2	1	niveaustabil oder -instabil

Tab. 10: Charakteristik der Experimentiertypen A – D

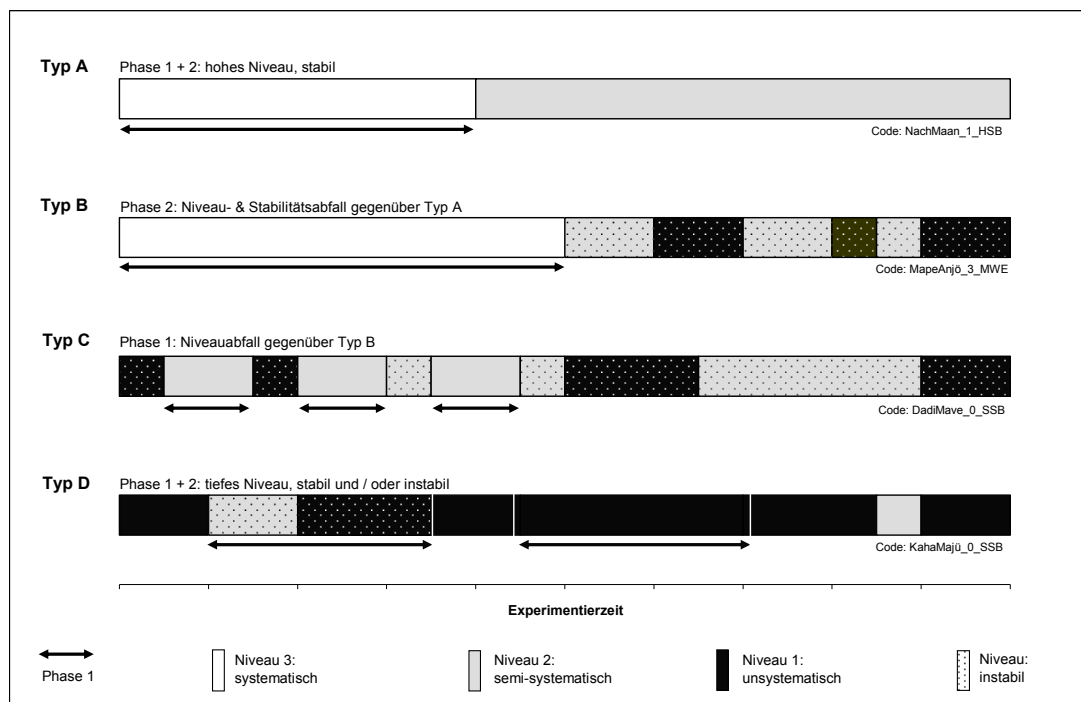


Abb. 10: Niveaustufengrafik von Repräsentantenpaaren der Experimentiertypen A – D

Verteilung der Experimentiertypen A – D in Anzahl der Probandenpaare					
Typ	KG	TG 1	TG 2	TG 3	Σ
A	2	10	3	5	20
B	2	–	1	4	7
C	4	1	–	1	6
D	4	1	1	2	8
Σ	12	12	5	12	41
Legende: KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3)					

Tab. 11: Verteilung der Experimentiertypen A – D

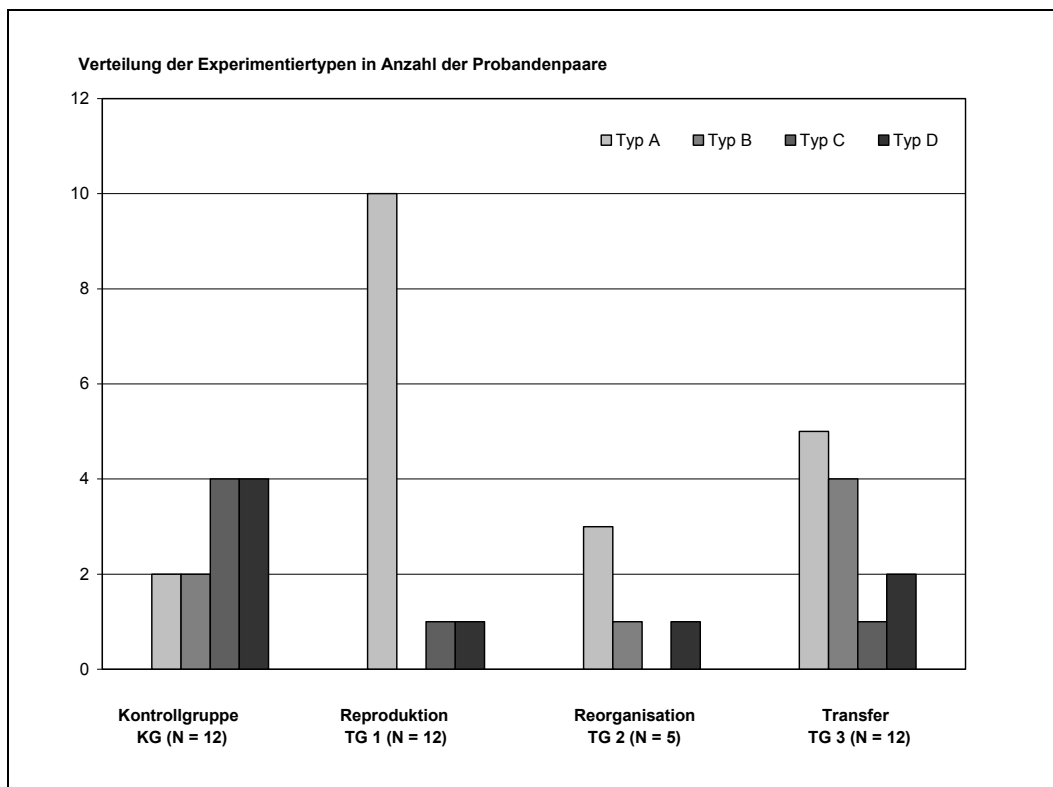


Abb. 11: Verteilung der Experimentiertypen A – D

Mit der Abbildung 10 wird die Verteilung der Niveaustufen von Repräsentantenpaaren der vier Experimentiertypen (A – D) über die ersten 100 Minuten der Experimentierzeit dargestellt. Die charakteristische Komposition aus den Kriterien Niveaustufe und Niveaustabilität der Typen A – D kann der Tabelle 10 entnommen werden. Mit der Tabelle 11 sowie der Abbildung 11 wird die Verteilung der Probandenpaare über die Experimentiertypen illustriert.

Es sei darauf verwiesen, dass die Prozessgrafiken (vgl. Kap. 4.4.1) und Niveaustufengrafiken nicht in einem gegenseitig ersetzenden, sondern in einem zueinander ergänzenden Verhältnis stehen. Der Vorteil der Prozessgrafiken (z. B. Abbildung 8 a und b) liegt in der Darstellung prozessbezogener Daten des Experimentierens. Der Vorteil der Niveaustufengrafiken (z. B. Abbildung 10) liegt in der Darstellung von Experimentierprozessen, die einer kriteriengeleiteten Bewertung unterzogen wurden.

Experimentiertyp A. Probanden, deren Experimentierverhalten nach dem Experimentiermuster A abläuft, zeichnen sich durch eine hohe Niveaustabilität und eine hohe Niveaustufe in beiden Phasen des Experimentierens aus. Ein Repräsentantenpaar dieses Experimentiertyps ist in der Abbildung 10 dargestellt. Deren hohe Niveaustabilität in beiden Phasen des Experimentierens wird durch das konstante, nicht wechselhafte Farbschema symbolisiert, wohingegen deren Niveaustufen durch die Farbauswahl gekennzeichnet werden. Das Repräsentantenpaar des Experimentiertyps besetzt in der Phase 1 die Niveaustufe 3 (systematisch) und in der Phase 2 die Niveaustufe 2 (semi-systematisch). Dieser beim Übergang von Phase 1 zu Phase 2 identifizierte Niveauabfall wird in der Niveaustufengrafik (Abbildung 10) durch den Farbwechsel symbolisiert. Trotz des auffälligen Niveauverlustes in Phase 2, ist das über beide Phasen erzielte Gesamtniveau des Experimentiertyps A – gemessen an der Gesamtheit aller Probanden aus der Teilstichprobe – als hoch einzustufen.

Von den 41 Probandenpaaren der Teilstichprobe folgen 20 Probandengruppen (49 %) dem Experimentierverlauf des Typs A, deren Verteilung über die Test- und Kontrollgruppen der Tabelle 11 sowie der Abbildung 11 zu entnehmen ist. Von den 20 Probandenpaaren des Experimentiertyps A weisen 19 Probandengruppen eine der Abbildung analoge Niveaustufenverteilung auf (Niveaustufe 3 in Experimentierphase 1, Niveaustufe 2 in Experimentierphase 2). Nur ein Probandenpaar der Testgruppen mit der vorbereitenden Instruktion Reproduktion (TG 1) experimentiert in beiden Phasen auf demselben, hohen Niveau der Stufe 3 (systematisch) ohne Niveaustufenwechsel.

In Abweichung zum Experimentierverlauf des dargestellten Repräsentantenpaares wurde bei fünf der 20 Probandenpaare die Phase 2 auch als stabil gewertet, wenn ein zeitlich begrenzter Niveaustufenwechsel von der semi-systematischen Niveaustufe (Niveau 2) auf die unterste Niveaustufe (Niveau 1) vorlag. Das Kriterium Niveaustabilität galt als erfüllt, wenn erstens der Niveaustufenwechsel nicht wiederholt

auftrat, zweitens die Experimentiersequenz auf der unteren Niveaustufe 10 – 15 % der Experimentierzeit nicht überstieg und drittens, wenn innerhalb einer Experimentiersequenz nicht wiederholt Experimentiertätigkeiten auf der untersten Niveaustufe durchgeführt wurden (z. B. mehrfache, konfundierende Änderungen an Profilkonstruktionen). Die häufigste Ursache des Niveaustufenwechsels war der einmalige, verführerische Versuch der Probandenpaare, zwei verschiedene Eigenkonstruktionen zu einer neuen, konfundierenden Profilkonstruktion zu kombinieren, verbunden mit dem Wunsch, deren Wirkung zu testen. Trotz des daraus resultierenden sequenziellen Niveaustufenwechsels, wurde dieses Experimentierverhalten als legitim gewertet und unter der Kategorie „Niveaustabilität“ subsumiert.

Der gegenteilige Effekt, ein zeitlich begrenzter, sequenzieller Wechsel von der semi-systematischen Niveaustufe (Niveau 2) auf die systematische Niveaustufe (Niveau 3) konnte bei keinem Vertreter des Experimentiertyps A festgestellt werden. Obwohl bei mehreren Probandenpaaren in der Phase 2 die Voraussetzungen zur Eingruppierung in die höchste Niveaustufe vorlagen („Aus eigenen Profilkonstruktionen nicht-konfundierende Messfolgen erstellen und aufnehmen“, vgl. Tätigkeiten des Experimentierens, Kap. 4.4.2), genügte die Datenauswertung der Probandenpaare nicht den Anforderungen der dritten Niveaustufe. Insbesondere das Kriterium „Vergleich und Bewertung einer bzw. mehrerer Messfolgen“ (vgl. V₃ bzw. V₄, Kap. 4.2.3 und Kap. 4.4.2) wurde nicht erfüllt. Das Missverhältnis von der Bereitstellung der Messdaten mit dem Potential zur niveauvollen Datenauswertung und der tatsächlich erfolgten Auswertung der Messdaten durch „paarweise Vergleiche mit dem Schluss auf die bessere Flügeleigenschaft“ (vgl. V₂, Kap. 4.4.2) war die Hauptursache dafür, warum noch nicht einmal kurzzeitige Anstiege von der semi-systematischen auf die systematische Niveaustufe in der Phase 2 zu verzeichnen waren.

Experimentiertyp B. Das Experimentierverhalten des Typs B unterscheidet sich vom Experimentierverhalten des Typs A durch einen Niveau- und Stabilitätsabfall in der Phase 2. Da die Phase 1 von Änderungen gegenüber Experimentiertyp A unberührt bleibt, bezieht sich die weitere Dokumentation des Experimentierverhaltens nur auf die Phase der Profilkonstruktion (Phase 2). Ein Repräsentantenpaar des Experimentiertyps ist in der Abbildung 10 aufgenommen worden.

Die wechselhafte Farbfolge zwischen den Niveaustufen symbolisiert die Niveaunstabilität, die durch die punktierte Markierung verstärkt dargestellt werden soll. Der Abbildung ist zu entnehmen, dass in der Phase 2 Sequenzen des unsystematischen Experimentierens (Niveaustufe 1) durch Sequenzen des semi-systematischen Experimentierens (Niveaustufe 2) zeitlich ausbalanciert werden. Wegen der immer wiederkehrenden, konfundierenden Arbeitsweise und/ oder der nicht messdatenbasierten Profilkonstruktionen und/ oder der Vergleiche mit Bewertungen von konfundierenden Messfolgen, ist die Experimentiertätigkeit

insgesamt für eine Eingruppierung in die semi-systematische Niveaustufe nicht ausreichend. Wenn wie hier, ein Wechsel zwischen zwei Experimentierniveaus vorliegt, dann wird die gesamte Phase zusammengefasst und bewertet. Die Bewertung richtet sich nach der unteren der beiden Niveaustufen. Das heißt, der beim Übergang von Phase 1 zu Phase 2 identifizierte Niveauverlust des Experimentiertyps A tritt beim Typ B mit Verstärkung auf.

Von den 41 Probandenpaaren der Teilstichprobe folgen sieben Probandenpaare (17 %) dem Experimentiermuster B. Eine Verteilung über die Test- und Kontrollgruppen ist der Tabelle 11 sowie der Abbildung 11 zu entnehmen. Von den sieben Probandenpaaren des vorgestellten Experimentiertyps weisen alle eine der Abbildung analoge Niveaustufenverteilung auf (Niveaustufe 3 in Experimentierphase 1, Niveaustufe 1 in Experimentierphase 2). Unterschiede bestehen jedoch in der Häufigkeit der Niveaustufenwechsel. Probandengruppen mit häufigen Niveaustufenwechseln sind überwiegend bei der Kontrollgruppe zu finden. Im Extremfall konnte einmalig ein Probandenpaar identifiziert werden, dass über eine nicht zu vernachlässigende Experimentierzeit, Niveaustufenwechsel im Zeittakt von fünf Minuten oder weniger vollzog. Das Probandenpaar verfolgte das Ziel, eine Auswahl von Profilen mit vergleichbarer Drehzahl durch Änderungen an den selbst hergestellten Profilkonstruktionen zu optimieren. Die schnelle Folge von konfundierenden und nicht-konfundierenden Experimenten verdeutlicht, dass es bei dieser Variante der Trial- & Error-Strategie nicht um die systematische Bestimmung abhängiger Variablen ging. Dieses beim Experimentiertyp B nur einmal festgestellte Experimentierverhalten ist beim Experimentiertyp A nicht und bei den Typen C und D mit größerer Häufigkeit zu finden. Demgegenüber sind Probandengruppen des Typs B mit wenigen Niveaustufenwechseln tendenziell eher in den Testgruppen zu finden. Bei der Bewertung wurde sichergestellt, dass die Ursache für den geringen Niveaustufenwechsel nicht auf einen Mangel an Experimentiermöglichkeiten, wie z. B. durch zeitlich ausgedehnte Eigenkonstruktionen mit wenigen Änderungen an den Profilkonstruktionen, zurückgeführt werden kann.

Experimentiertyp C. Das Experimentierverhalten des Typs C unterscheidet sich von dem in der Hierarchie vorangehenden Experimentiertypen B durch einen Niveauabfall in Phase 1. Kennzeichnend für diesen Experimentiertyp ist die niveaustabile, semi-systematische Experimentiertätigkeit (Niveaustufe 2) in Phase 1 und die niveauinstabile, unsystematische Experimentiertätigkeit (Niveaustufe 1) in Phase 2. Des Weiteren können bei den Probandenpaaren des Typs C über die experimentiertypenspezifischen Kriterien hinaus vier Effekte mit einer Prägnanz festgestellt werden, die bei den Experimentiertypen A und B nicht aufgetreten sind.

Ein Repräsentantenpaar des Experimentiertyps ist in der Abbildung 10 zu finden. Hier symbolisiert das konstante, nicht wechselnde Farbschema in Phase 1 das typenspezifische Experimentierverhalten. Anders als bei den vorangehenden

Vertretern der Typen A und B, weist die Repräsentantengruppe des Typs C in Phase 1 ein Experimentierverhalten mit Unterbrechungen durch Phase 2 auf (Effekt 1: Experimentiertätigkeiten in Phase 1 mit Unterbrechungen durch Phase 2). Das bedeutet, dass die Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelset nur sequenziell stattfindet, was dem Probandenpaar einen umfassenden Überblick über die messdatenbasierten Informationen erschwert. Ferner ist der Abbildung zu entnehmen, dass der Vertreter des Experimentiertyps mit einer nichtmessdatenbasierten Profilkonstruktion (Niveaustufe 1 in Experimentierphase 2) in die Experimentiertätigkeit einsteigt (Effekt 2: Einstieg in Experimentiertätigkeit mit Phase 2). Erst durch die fortschreitende Arbeit mit dem aufgabenspezifischen Experimentiermaterial Flügelsets (Phase 1, Niveaustufe 2), gelingt es dem Repräsentantenpaar, messdatenbasierte Profile zu konstruieren und/ oder messdatenbasierte Profiländerungen vorzunehmen. Entnommen werden kann dies der Niveaustufengrafik, die während bzw. nach Abschluss der Arbeit mit den Flügelsets für die Phase der Profilkonstruktion einen Niveauanstieg dokumentiert. Dieser auf die Profilkonstruktion nur kurzzeitig und nicht nachhaltig wirkende Einfluss der messdatenbasierten Arbeit mit den Flügelsets ist ein eigenständiger Effekt (Effekt 3: Einfluss der Experimentiertätigkeiten der Phase 1 auf eigene Profilkonstruktionen nicht nachhaltig), der erst nach Abschluss der Arbeit mit den Flügelsets durch eine Folge von Profilkonstruktionen auf wechselnden Niveaustufen identifiziert und beschrieben werden kann. Die mit zahlreichen Niveaustufenwechseln behaftete Experimentiertätigkeit des Repräsentantenpaares ist der Grund, warum die gesamte Phase 2 nur als instabil bewertet werden kann.

Es sei darauf verwiesen, dass der vierte und letzte Effekt (Effekt 4: Experimentiertätigkeiten mit extrem schnellen Niveaustufenwechseln) nicht durch das Repräsentantenpaar des Typs C in der Abbildung 10 wiedergegeben wird. Der Effekt bezieht sich auf den Zeittakt der Niveaustufenwechsel. Charakteristisch für den Effekt, der vorwiegend aber nicht ausschließlich in Phase 2 auftritt, sind extrem schnelle Folgen von Niveaustufenwechseln, die Zeitintervalle von fünf Minuten nicht übersteigen. Dahinter verbirgt sich häufig eine Experimentiertätigkeit, die durch schnell vorzunehmende Änderungen an Profilkonstruktionen wie Kürzen, Verbiegen oder Schleifen usw. geprägt sind.

Von den 41 Probandenpaaren der Teilstichprobe folgen sechs Probandengruppen (15 %) dem Experimentiermuster. Eine Verteilung über die Test- und Kontrollgruppen ist der Tabelle 11 sowie der Abbildung 11 zu entnehmen. Eine Verteilung der festgestellten Effekte auf die sechs Probandenpaare des Experimentiertyps C wird in der Tabelle 12 differenziert dargestellt. Zur Orientierung ist die Verteilung der Probandenpaare des Experimentiertyps C von der Tabelle 11 erneut mit aufgenommen worden.

Verteilung der Effekte des Experimentiertyps C in Anzahl der Probandenpaare					
Effekt	KG	TG 1	TG 2	TG 3	Σ Effekte
1	2	–	–	1	3
2	2	–	–	–	2
3	4	1	–	–	5
4	1	–	–	–	1
Σ Effekte	9	1	–	1	11
Typ C	4	1	–	1	6
Legende: 1 Experimentiertätigkeiten in Phase 1 mit Unterbrechungen durch Phase 2 2 Einstieg in Experimentiertätigkeit mit Phase 2 3 Einfluss der Experimentiertätigkeiten von Phase 1 auf eigene Profilkonstruktionen (Phase 2) <i>nicht nachhaltig</i> 4 Experimentiertätigkeiten mit extrem schnellen Niveaustufenwechseln KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3)					

Tab. 12: Verteilung der Effekte des Experimentiertyps C

Wie der Tabelle 12 zu entnehmen ist, sind bei den sechs Probandenpaaren des Typs C die vier Effekte in der Summe 11mal festzustellen. Eine Effekthäufung kommt bei den vier Probandengruppen der Kontrollgruppe vor. Aus dem Vergleich der Probandenzahl mit der Anzahl der Effekte ist zu schließen, dass bei den Probanden der Kontrollgruppe mehrere Effekte gleichzeitig, also eine Effektkombination oder -sammlung vorhanden sein muss. Auf eine weitere tabellarische Spezifizierung der Effektkombinationen in der Kontrollgruppe wird wegen des geringen Stichprobenumfangs verzichtet. Zu erwähnen ist aber, dass bei keinem der vier Probandenpaare alle vier Effekte gleichzeitig auftreten, wohingegen bei der einen Hälfte der Kontrollgruppe zwei und bei der anderen Hälfte drei Effektkombinationen auftreten. Der Effekt 3 ist bei allen Probandenpaaren ein Element der Effektkombination. Hier liegt die Vermutung nahe, dass eine Ursache des Effekts in den Effekten 1 und 2 zu suchen ist.

Experimentiertyp D. Probandenpaare, deren Experimentierverhalten nach dem Experimentiermuster D abläuft, fallen durch eine niedrige Niveaustufe in beiden Phasen des Experimentierens auf. Aufgrund dessen wird eine differenzierte Darstellung zwischen niveaustabiler und niveauinstabiler Experimentiertätigkeit nicht mehr vorgenommen. In der Charakteristik des Experimentiertyps (vgl. Tabelle 10) fasst die Kennzeichnung „stabil oder instabil“ die Kombinationsmöglichkeiten zusammen, die als stabil-unsystematisches oder als instabil-unsystematisches Experimentieren (d. h. mit sequenziellen Niveaustufenwechseln auf eine höhere Stufe) zu interpretieren sind. Der Tabelle ist ferner zu entnehmen, dass in der Charakteristik des Experimentiertyps auch Probandenpaare mit aufgenommen werden, die in der Phase 1 die Niveaustufe 2 erzielen. Sie grenzen sich jedoch durch ihre instabile Experimentiertätigkeit (d. h. mit sequenziellen Niveaustufenwechseln auf eine andere Niveaustufe) von Probandenpaaren mit identischer Niveaustufenverteilung des Typs C ab. Trotz der semi-systematischen Experimentiertätigkeit in Phase 1, ist das Gesamtniveau dieser Probandenpaare wegen des niveauinstabilen Verlaufs der Experimentierphase als niedrig einzuordnen.

Von den 41 Probandenpaaren der Teilstichprobe folgen 8 Probandenpaare (20 %) dem Experimentiermuster. Eine Verteilung über die Test- und Kontrollgruppen ist der Tabelle 11 sowie der Abbildung 11 zu entnehmen. Von den 8 Probandenpaaren erreichen etwa ein Drittel in der Phase 1 die Niveaustufe 2. Etwa zwei Drittel der Probandenpaare experimentieren in beiden Phasen auf der Niveaustufe 1 (vgl. Repräsentantenpaar Abbildung 10). In Analogie zum Experimentiertyp C können beim Typ D über die experimentiertypenspezifischen Kriterien hinaus dieselben vier Effekte identifiziert werden, die die Experimentiertätigkeit der Probandenpaare begleiten. Eine differenzierte Verteilung der Effekte über die Test- und Kontrollgruppen auf die 8 Probanden des Typs D ist der Tabelle 13 zu entnehmen. Zur Orientierung ist die Verteilung der Probandenpaare des Experimentiertyps von der Tabelle 11 übertragen worden.

Wie durch die Übersicht mit der Tabelle 13 vermittelt wird, sind bei den 8 Probandenpaaren des Typs D über die typenspezifische Kennung hinaus, die Effekte 1 – 4 in der Summe 13mal festzustellen. Mit derselben Prägnanz wie beim Experimentiertyp C ist eine Effekthäufung bei den vier Probandengruppen der Kontrollgruppe zu erkennen. Aus dem Vergleich der Probandenpaarzahlen mit der Anzahl der Effekte ist zu schließen, dass Effektkombinationen auftreten müssen. Aufgrund des geringen Stichprobenumfangs wird auf eine weitere Spezifizierung der Effektkombinationen verzichtet. Zu erwähnen ist, dass bei keinem der Probandenpaare alle vier Effekte gleichzeitig auftreten. Der Effekt 3 ist bei allen Probandenpaaren ein Element der Effektkombination. Im Unterschied zum Experimentiertyp C, bei dem der Einfluss der Arbeit mit den Flügelsets auf die eigenen Profilkonstruktionen (Effekt 3) wenig nachhaltig ist, ist der Einfluss der

Arbeit mit den Flügelsets auf die eigenen Profilkonstruktionen bei den Probandenpaaren des Experimentiertyps D tendenziell eher wirkungslos. Dies ist vor allem dann festzustellen, wenn die Experimentierphase 1 mit der tiefsten Niveaustufe bewertet werden muss.

Verteilung der Effekte des Experimentiertyps D in Anzahl der Probandenpaare					
Effekt	KG	TG 1	TG 2	TG 3	Σ Effekte
1	2	–	–	1	3
2	2	–	–	–	2
3	4	1	1	1	7
4	1	–	–	–	1
Σ Effekte	9	1	1	2	13
Typ D	4	1	1	2	8
Legende: 1 Experimentiertätigkeiten in Phase 1 mit Unterbrechungen durch Phase 2 2 Einstieg in Experimentiertätigkeit mit Phase 2 3 Einfluss der Experimentiertätigkeiten von Phase 1 auf eigene Profilkonstruktionen (Phase 2) <i>wirkungslos</i> 4 Experimentiertätigkeiten mit extrem schnellen Niveaustufenwechseln KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3)					

Tab. 13: Verteilung der Effekte des Experimentiertyps D

4.4.4 Niveaustufen des Experimentierens

Durch die Niveaustufen des Experimentierens werden die Fähigkeiten der Probandenpaare zum Transfer des Variablenkonzeptes von einer vorbereitenden Instruktion auf die offene Experimentieraufgabe dokumentiert und zwar

- a) in Abhängigkeit eines Kontextwechsels und
- b) in Abhängigkeit eines Kontextwechsels unter der Bedingung, dass zusätzlich ein Transfer von der Arbeit mit vorgegebenem Experimentiermaterial (Experi-

mentierphase 1) auf die Arbeit mit nicht vorgegebenem Experimentiermaterial (Experimentierphase 2) erbracht werden muss.

Der zweite, zusätzliche Transfer, der durch Änderung der Experimentierbedingungen geleistet werden muss (vgl. b), wird durch den Übergang von der Experimentierphase 1 auf die Experimentierphase 2 abgebildet, da in der vorbereitenden Instruktion das Konzept zum Umgang mit Variablen nur mit Hilfe von vorgegebenem Experimentiermaterial vermittelt wird.

Damit aus den Beobachtungsdaten beide Transferleistungen differenziert analysiert werden können, wird in der Abbildung 12 die Verteilung der Niveaustufen des Experimentierens nicht nur in Abhängigkeit der Kontexte der Vorbereitung, sondern auch in Abhängigkeit der beiden Experimentierphasen dargestellt. Um die Graduierung der Kontextunterscheide zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe anzuzeigen, werden die Begriffe Reproduktion (gleiche Kontexte, TG 1), Reorganisation (ähnliche Kontexte, TG 2) und Transfer (verschiedene Kontexte, TG 3) verwendet. Ergänzt wird die Abbildung durch die Tabelle 14, so dass nicht nur für jede Experimentierphase getrennt, sondern für beide Experimentierphasen zusammen, die Verteilung der Niveaustufen (Niveaustufenkombination) abgelesen werden kann.

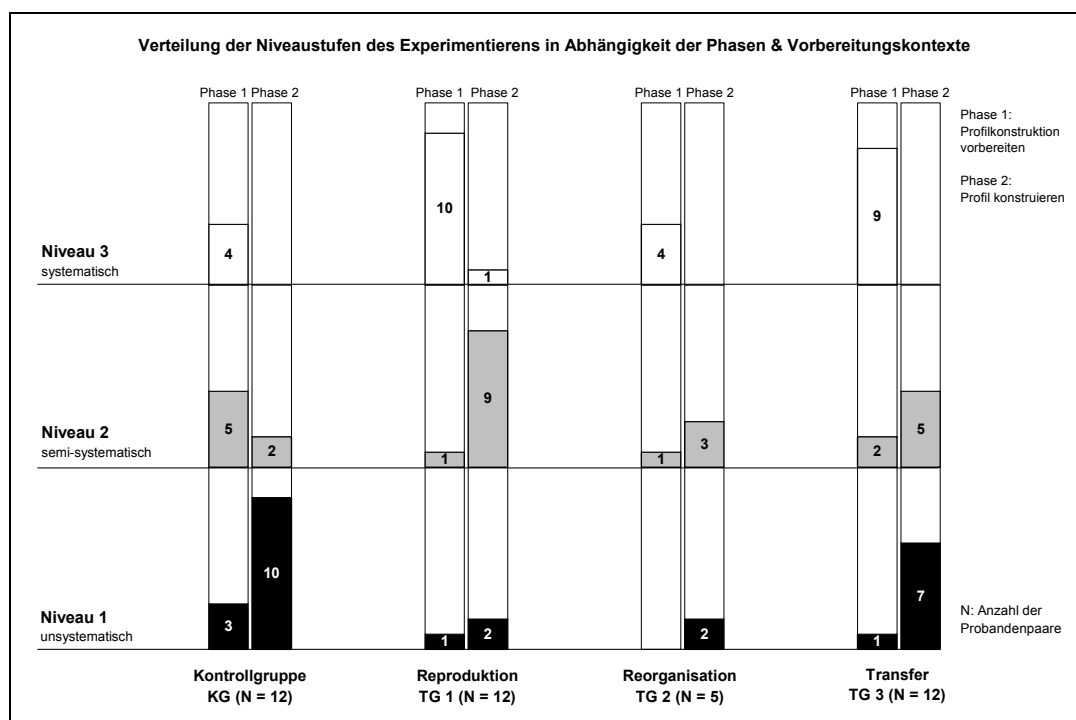


Abb. 12: Verteilung der Niveaustufen des Experimentierens in Abhängigkeit der Experimentierphasen und Kontexte der vorbereitenden Instruktion

Abbildung 12 zeigt die Besetzung der Niveaustufen durch die Anzahl der Probandenpaare. Zur Auswertung steht für jede Test- und Kontrollgruppe mit $N =$

12 dieselbe Anzahl an Probandenpaaren zur Verfügung. Davon weicht mit $N = 5$ die Anzahl der Probandenpaare der Testgruppe 2 (Reorganisation) ab. Wegen der geringen Probandenpaaranzahl ist die vergleichende, qualitative Diskussion mit den weiteren Test- und Kontrollgruppen (TG 1, TG 3, KG) nur mit Einschränkung gültig.

Ausgewertet wird die Verteilung der Niveaustufen, indem für die Test- und Kontrollgruppen paarweise Vergleiche durchgeführt werden, zunächst nur für die erste Experimentierphase (s. u.), anschließend nur für die zweite Experimentierphase (s. u.) und abschließend für beide Experimentierphasen (s. u.). Der vergleichenden Diskussion folgen statistische Auswertungsverfahren, so dass die Ergebnisse für die Niveaustufen des Experimentierens quantitativ gestützt werden.

Qualitative Auswertung der Niveaustufen des Experimentierens. Qualitativer Vergleich der ersten Experimentierphase. In der Testgruppe Reproduktion (gleiche Kontexte, TG 1) erreichen fast alle Probandenpaare die höchste Niveaustufe des Experimentierens (Niveaustufe 3). Mit $N = 10$ Probandenpaare sind das mehr als doppelt so viele Probandenpaare als in der Kontrollgruppe (ohne Vorbereitung, KG). Die Kontrollgruppe zeichnet sich durch eine annähernd gleichverteilte Besetzung der drei Niveaustufen aus. Von der Testgruppe Reproduktion (gleiche Kontexte, TG 1) gibt es gegenüber der Testgruppe Transfer (verschiedene Kontexte, TG 3), die zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe den größten Kontexttransfer leisten muss, nur geringe Abweichungen. Mit der Besetzung der Niveaustufe 3 durch $N = 9$ Probandenpaare ist der Kontrast zur Kontrollgruppe ähnlich hoch wie der Kontrast der Testgruppe Reproduktion (TG 1) zur Kontrollgruppe. Ebenfalls sehr vielversprechend ist das Ergebnis der Testgruppe Reorganisation (ähnliche Kontexte, TG 2), obwohl die Anzahl der Probandenpaare von $N = 5$ auf $N = 12$ nicht hochtransformiert werden darf: vier der fünf Probandenpaare erreichen die höchste Niveaustufe.

Durch einen Vergleich aller drei Testgruppen kann für die erste Experimentierphase festgestellt werden, dass fast alle Probandenpaare die höchste Niveaustufe besetzen, so dass zwischen den Testgruppen Reproduktion (TG 1), Reorganisation (TG 2) und Transfer (TG 3) nur geringe Unterschiede in der Verteilung der Probandenpaare über die Niveaustufen auftreten. Wenn in den Vergleich die Kontrollgruppe mit einbezogen wird, dann ist gegenüber jeder Testgruppe eine Verschiebung der Niveaustufenverteilung zu registrieren. In der Kontrollgruppe ist die Verschiebung der Niveaustufenverteilung an der höchsten Niveaustufe durch eine Abnahme und an den Niveaustufen zwei und drei durch eine Zunahme der Anzahl der Probandenpaare zu bemerken.

Zusammenfassung der qualitativen Ergebnisse der ersten Experimentierphase (Ergebnis 1). Als Ergebnis der umfassenden Vergleiche der ersten Experimentierphase kann festgestellt werden, dass es den meisten Probandenpaaren ohne

vorbereitende Instruktion (Kontrollgruppe) beim Experimentieren nicht gelingt, mit Variablen systematisch umzugehen. Demgegenüber kann die Wirkung einer einmaligen, einführenden, informierenden, expliziten und vorbereitenden Instruktion auf die Tätigkeit des Experimentierens an der Verteilung der Niveaustufen der Testgruppen belegt werden: Den meisten Probandenpaaren gelingt beim Experimentieren der systematische Umgang mit Variablen auf der höchsten Niveaustufe. Die Probandenpaare zeigen die Fähigkeit zum Transfer eines Variablenkonzeptes von einer vorbereitenden Instruktion auf eine offene Experimentieraufgabe. Weil zudem zwischen den Niveaustufenverteilungen der drei Testgruppen nur wenige Unterschiede auftreten, kann als ein weiteres Ergebnis konstatiert werden, dass die Transferleistungen (Reproduktion, Reorganisation, Transfer) vom Grad der Kontextunterschiede unabhängig sind.

Qualitativer Vergleich der zweiten Experimentierphase. Wie der Abbildung 12 zu entnehmen ist, wird in der zweiten Experimentierphase die höchste Niveaustufe nur von einem der 41 Probandenpaare erreicht. Die Diskussion über die Verteilung der Niveaustufen konzentriert sich deshalb auf das Besetzungsverhältnis zwischen der semi-systematischen und der unsystematischen Niveaustufe. In der Testgruppe Reproduktion (gleiche Kontexte, TG 1) erzielen 10 von 12 Probandenpaaren ein Experimentierniveau oberhalb der unsystematischen Stufe (Niveaustufe 1). Demgegenüber besetzen in der Kontrollgruppe (ohne Vorbereitung, KG) 10 von 12 Probandenpaaren ein Experimentierniveau unterhalb der semi-systematischen Stufe (Niveaustufe 2). Zwischen den Niveaustufenverteilungen der Testgruppe Reproduktion und der Kontrollgruppe liegen die Niveaustufenverteilungen der weiteren Testgruppen. Ein paarweiser Vergleich der Testgruppen Reproduktion (gleiche Kontexte, TG 1) mit der Testgruppe Transfer (verschiedene Kontexte, TG 3) dokumentiert, dass mit Zunahme des Kontexttransfers die Anzahl der Probandenpaare, die die Niveaustufe zwei besetzen stagniert und die Anzahl der Probandenpaare, die die Niveaustufe eins besetzen wächst. Weiterhin zeigt eine Gegenüberstellung der Ergebnisse der Testgruppe Transfer (TG 3) und der Kontrollgruppe (KG), dass die Tendenz des Besetzungsverhältnisses zwischen der semi-systematischen und der unsystematischen Niveaustufe fortgesetzt wird. Folglich ist der Kontrast der Testgruppe Transfer zur Kontrollgruppe weniger deutlich als der Kontrast der Testgruppe Reproduktion zur Kontrollgruppe. Über eine exakte Positionierung der Testgruppe Reorganisation (ähnliche Kontexte, TG 2) in der nach abnehmender Niveaustufenverteilung sortierten Folge „Testgruppe Reproduktion (TG 1), Testgruppe Transfer (TG 3), Kontrollgruppe (KG)“ kann aufgrund der abweichenden, ungleichen Probandenpaaranzahl der Testgruppe Reorganisation (TG 2) mit qualitativen Auswertungsverfahren nicht entschieden werden.

Zusammenfassung der qualitativen Ergebnisse der zweiten Experimentierphase (Ergebnis 2). Im Gegensatz zum Ergebnis 1, muss bei der Ergebnisformulierung der qualitativen Auswertung für den Vergleich der Niveaustufen der zweiten Experimentierphase bedacht werden, dass die Probandenpaare gleichzeitig einen Kontexttransfer und einen Transfer von der Arbeit mit vorbereitetem Experimentiermaterial (Phase 1) auf die eigene Profilkonstruktion (Phase 2) leisten müssen.

Als erstes Ergebnis kann durch den Vergleich der Testgruppe Reproduktion (TG 1) mit der Kontrollgruppe (KG) die Wirkung der einmaligen, einführenden, informierenden, expliziten und vorbereitenden Instruktion auf die Tätigkeit des Experimentierens festgestellt werden: Fast alle Probandenpaare, die keine vorbereitende Instruktion erhalten haben (Kontrollgruppe), können beim Experimentieren in der Phase 2 nur unsystematisch mit Variablen umgehen. Als zweites Ergebnis der Niveaustufenvergleiche kann festgestellt werden, dass die Fähigkeit der Probandenpaare zum Transfer eines Variablenkonzeptes von einer vorbereitenden Instruktion mit vorgegebenem Experimentiermaterial auf die eigene Profilkonstruktion von der Graduierung der Kontextunterschiede zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe abhängt. Die Kontextabhängigkeit kann an der Anzahl der Probandenpaare abgelesen werden, die das semi-systematische Experimentierniveau erreichen: In der Testgruppe Reproduktion (gleiche Kontexte, TG 1) erreichen fast doppelt so viele Probandenpaare die Niveaustufe zwei wie in der Testgruppe Transfer (verschiedene Kontexte, TG 3) und in der Testgruppe Transfer (TG 3) erreichen mehr als doppelt so viele Probandenpaare das semi-systematische Experimentierniveau wie in der Kontrollgruppe (KG).

Qualitativer Vergleich der ersten und zweiten Experimentierphase (Niveaustufenkombination). Der Abbildung 12 können für die Test- und Kontrollgruppen keine Informationen über die Verteilung der Niveaustufenkombinationen beider Experimentierphasen entnommen werden. Damit beim Vergleich der Niveaustufenverteilungen Fehlinterpretationen vermieden werden, wird die Abbildung 12 durch die Tabelle 14 ergänzt. Drei wichtige Informationen können der Tabelle entnommen werden. Information 1: Kein Probandenpaar erreicht in der Experimentierphase 2 eine höhere Niveaustufe als in der Experimentierphase 1, was bedeutet, dass eine Niveausteigerung beim Experimentieren durch einen Wechsel der Experimentierbedingungen von der Arbeit mit vorgegebenem Experimentiermaterial zur eigenen Profilkonstruktion nicht stattfindet. Information 2: Der Übergang von der Experimentierphase 1 zur Experimentierphase 2 ist für 35 der 41 Probandenpaare mit einem Niveauabfall (Niveaugap) verbunden. Der Effekt ist für alle Test- und Kontrollgruppen höchst charakteristisch, weil in jeder Gruppe für mindestens 9 der 12 Probandenpaare bzw. für alle fünf Probandenpaare der Testgruppe

Reorganisation (TG 2) ein Niveauabfall nachgewiesen werden kann. Obwohl zwischen der Niveaustufenkombinationen 3 und 2 (Synonyme Schreibweise: 3/ 2) sowie 2 und 1 (Synonyme Schreibweise: 2/ 1) die Niveaudifferenzen identisch sind, ist es möglich, eine Folge nach steigendem Niveauverlust zu erstellen, weil die Niveaustufenkombination 3/ 2 sowohl das höhere Ausgangsniveau (Experimentierphase 1) als auch das höhere Endniveau (Experimentierphase 2) hat. Dies erlaubt den Nachweis, dass die Stärke des Niveauverlustes von der Gruppenzugehörigkeit der Probandenpaare abhängig ist. Information 3: Sechs der 41 Probandenpaare besetzen für beide Experimentierphasen identische Niveaustufen. Davon besetzt nur ein Probandenpaar die systematische Niveaustufenkombination 3/ 3 und fünf Probandenpaare besetzen die unsystematische Niveaustufenkombination 1/ 1. Die semi-systematische Niveaustufe mit der Niveaustufenkombination 2/ 2 wird von keinem Probandenpaar belegt.

Verteilung der Niveaustufenkombinationen in Anzahl der Probandenpaare (N)						
Niveaustufen		KG	TG 1	TG 2	TG 3	Σ_N
Phase 1	Phase 2					
3	3	–	1	–	–	1
3	2	2	9	3	5	19
3	1	2	–	1	4	7
2	2	–	–	–	–	–
2	1	5	1	1	2	9
1	1	3	1	–	1	5
Σ_N		12	12	5	12	41

Tab. 14: Verteilung der Niveaustufenkombinationen

Mit den Informationen der Tabelle 14 kann die Verteilung der Niveaustufenkombinationen für beide Experimentierphasen analysiert werden. Der Fokus der Analyse liegt auf der Dokumentation des Niveauabfalls.

In der Testgruppe Reproduktion (gleiche Kontexte, TG 1) liegt beim Übergang von der Experimentierphase 1 zur Experimentierphase 2 für 9 von 12 Probandenpaaren ein Niveauabfall mit der Niveaustufenkombination 3/ 2 vor. Die Niveaudifferenz besagt, dass es den Probandenpaaren in Phase 2 nicht gelingt, das systematische Experimentierniveau zu erreichen, wenn sie zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe identische Kontexte vorfinden und einen Transfer von der Arbeit mit vorgegebenem Experimentiermaterial auf die eigene, freie Profilkonstruktion leisten müssen. Weiterhin ist für die Testgruppe kennzeichnend, dass die Niveaustufenkombination 3/ 1 von keinem Probandenpaar besetzt wird. Obwohl in der Kontrollgruppe (KG) ein Niveauabfall bei einer vergleichbaren Anzahl an Probandenpaaren vorkommt, ist gegenüber der Testgruppe Reproduktion (TG 1) eine markante Zunahme des Niveaugaps zu registrieren. Die Verstärkung des Effekts kann an der Abnahme der Probandenpaaranzahl mit der Niveaustufenkombination 3/ 2, der geringfügigen Zunahme der Probandenpaaranzahl mit der Niveaustufenkombination 3/ 1 sowie der starken Zunahme der Probandenpaaranzahl mit der Niveaustufenkombination 2/ 1 erkannt werden. Die Niveaustufenkombination 2/ 1 wird zusammen mit der Kombination 1/ 1 von zwei Dritteln der Probandenpaare der Kontrollgruppe besetzt. Zwischen der Verteilung der Niveaustufenkombinationen der Testgruppe Reproduktion (TG 1) und der Kontrollgruppe (KG) können die Verteilungen der Niveaustufenkombinationen der weiteren Testgruppen eingereiht werden. Das ergeben paarweise Vergleiche der Probandenpaaranzahlen zwischen der Testgruppe Reproduktion (gleiche Kontexte, TG 1) und der Testgruppe Transfer (verschiedene Kontexte, TG 3) sowie zwischen der Testgruppe Transfer (TG 3) und der Kontrollgruppe (ohne Vorbereitung, KG). Für den ersten Vergleich ist eine Stagnation der Niveaustufenkombinationen 3/ 2 und eine Zunahme der Niveaustufenkombination 3/ 1 zu dokumentieren. Für den zweiten Vergleich kann eine weitere Verschiebung von den Niveaustufenkombinationen 3/ 2 und 3/ 1 zu den Niveaustufenkombinationen 2/ 1 und 1/ 1 hin festgestellt werden. Die Verteilung der Niveaustufenkombinationen der Testgruppe Reorganisation (ähnliche Kontexte, TG 2) steht nicht im Widerspruch zu den Ergebnissen. In Analogie zur Niveaustufenverteilung der Experimentierphase 2 kann über eine exakte Positionierung der Testgruppe Reorganisation (TG 2) in der nach steigendem Niveauverlust sortierten Folge „Testgruppe Reproduktion (TG 1), Testgruppe Transfer (TG 3), Kontrollgruppe (KG)“ aufgrund der abweichenden, ungleichen Probandenpaaranzahl mit qualitativen Auswertungsverfahren nicht entschieden werden.

Zusammenfassung der qualitativen Ergebnisse der ersten und zweiten Experimentierphase (Ergebnis 3). Durch den Vergleich der Niveaustufen der ersten und zweiten Experimentierphase (Niveaustufenkombination) kann für die Test- und Kontrollgruppen ein Niveauverlust identifiziert werden, sofern die Probandenpaare

in der Phase 1 die unsystematische Niveaustufe übertreffen. Der Niveauverlust, d. h. die Abnahme der Fähigkeit zum Transfer eines Variablenkonzeptes kann auf die Änderung der Experimentierbedingungen zurückgeführt werden, die durch den Wechsel von der Arbeit mit vorbereitetem Experimentiermaterial (Phase 1) zu der Arbeit mit nicht vorbereitetem Experimentiermaterial (Phase 2) geschaffen wird.

Die Stärke des Niveauverlustes variiert mit dem Grad der Kontextunterschiede zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe. Die Probandenpaare der Testgruppe Transfer (verschiedene Kontexte, TG 3) zeigen einen größeren Niveauabfall als die Probandenpaare der Testgruppe Reproduktion (gleiche Kontexte, TG 1). Wenn die Probandenpaare keine vorbereitende Instruktion (Kontrollgruppe) erhalten, dann ist der Niveauverlust am stärksten ausgeprägt.

Quantitative Auswertung der Niveaustufen des Experimentierens. Die drei Ergebnisse der qualitativen Auswertung sollen durch statistische Analysen gestützt werden. Der Mann-Whitney U-Test wird für den Vergleich der Ergebnisse der ersten Experimentierphase (Ergebnis 1) und für den Vergleich der Ergebnisse der zweiten Experimentierphase (Ergebnis 2) eingesetzt. Der U-Test ist das nicht-parametrische Pendant zum t-Test für unabhängige Stichproben, der für ein Zwei-Gruppen-Experiment (paarweiser Vergleich) eine Signifikanzprüfung durch Vergleich von Rangsummen ausführt. Mit dem Wilcoxon-Test werden die Ergebnisse der Niveaustufenkombinationen (Ergebnis 3) auf Signifikanz geprüft. Der Test ist das nichtparametrische Pendant zum t-Test für zwei abhängige Stichproben.

Die statistischen Datenanalysen können nicht mit t-Tests durchgeführt werden, weil die Niveaustufen des Experimentierens keine Intervallskalenqualität haben. Die Niveaustufen des Experimentierens sind rangskalierte Daten, deren Rangplätze von mehreren Probandenpaaren geteilt werden (Rangbindung bzw. verbundene Ränge). Die Probandenpaare verteilen sich in der ersten Experimentierphase auf drei Rangplätze und in der zweiten Experimentierphase auf zwei Rangplätze. Für beide Testverfahren sind die mathematischen Voraussetzungen erfüllt, dass die Versuchspersonen (Probandenpaare) zufällig den Test- und Kontrollgruppen zugeordnet werden. Weil der Mann-Whitney U-Test und der Wilcoxon-Test Signifikanzprüfungen für Probanden erstellen, wird die Anzahl der Probanden durch Verdopplung der Anzahl der Probandenpaare angegeben. Eine kritische Bewertung der Maßnahme folgt nach den Analysen. Für die Testverfahren wird mit den p-Werten die zweiseitige Signifikanz berechnet, weil die Ergebnisse einer ungerichteten Fragestellung (zweiseitige Signifikanzprüfung) weniger überschätzt und nicht so schnell signifikant werden wie die Ergebnisse einer gerichteten Fragestellung (einseitige Signifikanzprüfung).

Mann-Whitney U-Tests auf Gruppenunterschiede beim Experimentieren Vergleich der Experimentierphase 1						
Phase	Gruppe 1	Gruppe 2	N ₁	N ₂	Z _{korrr.}	p-Wert
1	TG 1	KG	24	24	3,25	0,001
1	TG 2	KG	10	24	2,53	0,012
1	TG 3	KG	24	24	2,78	0,005
1	TG 1	TG 2	24	10	0,09	0,932
1	TG 2	TG 3	10	24	0,38	0,702
1	TG 1	TG 3	24	24	0,63	0,531
Legende: KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3) N Anzahl der Probanden: Gruppe 1 (N ₁), Gruppe 2 (N ₂) Z _{korrr.} Z-Wert korrigiert um verbundene Ränge (Rangbindung)						

Tab. 15: Mann-Whitney U-Tests mit Rangbindung zwischen Test- und Kontrollgruppen für Experimentierphase 1. **Fettdruck:** signifikante Effekte für $p < 0,05$

Die Ergebnisse der Mann-Whitney U-Tests sind für $p < 0,05$ signifikant, wenn die Testgruppen paarweise mit der Kontrollgruppe verglichen werden. Die Ergebnisse sind nicht signifikant, wenn die Testgruppen untereinander verglichen werden. Die Interpretation der Daten stimmt mit dem Ergebnis 1 überein: Den Probanden der Kontrollgruppe ist es ohne vorbereitende Instruktion nicht möglich, die naturwissenschaftliche Arbeitsweise Umgang mit Variablen effektiv beim Experimentieren umzusetzen. Die Tätigkeit des Experimentierens ist dagegen für solche Probanden effektiv, die eine vorbereitende Instruktion erhalten haben. Weiterhin ist die Tätigkeit des Experimentierens für Probanden effektiv, die zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe einen Kontextwechsel leisten müssen. Der fachliche Kontext der vorbereitenden Instruktion hat beim Experimentieren keinen Einfluss auf die naturwissenschaftliche Arbeitsweise, wenn die Probanden mit vorgegebenem Experimentiermaterial arbeiten (Phase 1).

Mann-Whitney U-Tests auf Gruppenunterschiede beim Experimentieren Vergleich der Experimentierphase 2						
Phase	Gruppe 1	Gruppe 2	N ₁	N ₂	Z _{kor.}	p-Wert
2	TG 1	KG	24	24	4,57	< 0,001
2	TG 2	KG	10	24	2,47	0,014
2	TG 3	KG	24	24	1,87	0,061
2	TG 1	TG 2	24	10	1,59	0,112
2	TG 2	TG 3	10	24	0,94	0,347
2	TG 1	TG 3	24	24	3,09	0,002
Legende: KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3) N Anzahl der Probanden: Gruppe 1 (N ₁), Gruppe 2 (N ₂) Z _{kor.} Z-Wert korrigiert um verbundene Ränge (Rangbindung)						

Tab. 16: Mann-Whitney U-Tests mit Rangbindung zwischen Test- und Kontrollgruppen für Experimentierphase 2. **Fettdruck:** signifikante Effekte für $p < 0,05$

Die U-Tests führen für die zweite Experimentierphase nicht immer zu signifikanten Ergebnissen, wenn die Kontrollgruppe paarweise mit den Testgruppen verglichen wird. Signifikant sind die Unterschiede zwischen der Kontrollgruppe (KG) und der Testgruppe Reproduktion (TG 1) sowie zwischen der Kontrollgruppe (KG) und der Testgruppe Reorganisation (TG 2). Nicht signifikant ist der Unterschied zwischen der Kontrollgruppe (KG) und der Testgruppe Transfer (TG 3), die zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe den größten Kontexttransfer leisten muss.

Die signifikanten Ergebnisse erlauben die Interpretation, dass die Probanden die Fähigkeit zum Transfer eines Variablenkonzeptes von einer vorbereitenden Instruktion mit vorgegebenem Experimentiermaterial auf die eigene Profilkonstruktion haben. In Kombination mit dem nicht signifikanten Ergebnis wird die Abhängigkeit der Transferfähigkeit des Variablenkonzeptes vom Grad der Kontextunterscheide zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe deutlich. Den Probanden der Testgruppe Transfer (TG 3) gelingt der Transfer eines Variablenkonzeptes von einer vorbereitenden Instruktion mit vorgegebenem Experimentiermaterial auf die eigene Konstruktion von Windradprofilen nicht besser als den Probanden der Kontrollgruppe (KG), die keine vorbereitende

Instruktion erhalten haben. Das besagt der nicht signifikante Unterschied zwischen der Testgruppe Transfer (TG 3) und der Kontrollgruppe (KG). Die Ergebnisse werden durch die paarweisen Vergleiche der Testgruppen gestärkt. Dies kann aus dem nicht signifikanten Unterschied zwischen den Testgruppen Reproduktion (TG 1) und Reorganisation (TG 2), dem nicht signifikanten Unterschied zwischen den Testgruppen Reorganisation (TG 2) und Transfer (TG 3) und dem signifikanten Unterschied zwischen den Testgruppen Reproduktion (TG 1) und Transfer (TG 3) herausgelesen werden. Zusammen stützen die sechs U-Tests die qualitativen Ergebnisse für die Experimentierphase 2 (Ergebnis 2).

Mann-Whitney U-Tests auf Gruppenunterschiede beim Experimentieren Aufhebung der Phasentrennung zwischen Experimentierphase 1 und 2						
Phase	Gruppe 1	Gruppe 2	N₁	N₂	Z_{kor.}	p-Wert
o. U.	TG 1	KG	24	24	3,02	0,003
o. U.	TG 2	KG	10	24	1,89	0,059
o. U.	TG 3	KG	24	24	1,70	0,089
o. U.	TG 1	TG 2	24	10	0,43	0,665
o. U.	TG 2	TG 3	10	24	0,46	0,644
o. U.	TG 1	TG 3	24	24	1,17	0,242
Legende: o. U. ohne Unterscheidung, Aufhebung der Phasentrennung zwischen Experimentierphase 1 und 2 KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3) N Anzahl der Probanden: Gruppe 1 (N ₁), Gruppe 2 (N ₂) Z _{kor.} Z-Wert korrigiert um verbundene Ränge (Rangbindung)						

Tab. 17: Mann-Whitney U-Tests mit Rangbindung zwischen Test- und Kontrollgruppen mit Aufhebung der Experimentierphasentrennung zwischen Experimentierphase 1 und 2.

Fettdruck: signifikante Effekte für $p < 0,05$

Mit den vorausgehenden U-Tests soll eine Signifikanzprüfung vorgenommen werden, wenn die Unterscheidung zwischen den Experimentierphasen 1 und 2 aufgehoben wird (Tabelle 17). Das kann realisiert werden, indem beispielsweise die unsystematischen, semi-systematischen und systematischen Niveaustufen beider Experimentierphasen für jede Gruppe aufsummiert werden. Dieses methodische

Verfahren erlaubt eine Aussage über die Wirkung einer vorbereitenden Instruktion in Abhängigkeit der Kontextunterschiede zwischen Vorbereitung und Experimentieraufgabe, wenn die Transferleistung von der Arbeit mit vorgegebenem Experimentiermaterial (Phase 1) auf die Arbeit mit nicht vorgegebenem Experimentiermaterial (Phase 2) vernachlässigt wird.

Unter Aufhebung der Experimentierphasentrennung kann nur für den Vergleich der Testgruppe Reproduktion (TG 1) mit der Kontrollgruppe (KG) ein signifikanter Unterschied festgestellt werden. Die Ergebnisse der U-Tests führen zu der Interpretation, dass nur die Probanden eine Fähigkeit zum Transfer des Variablenkonzeptes zeigen, die zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe identische Kontexte vorfinden (Tabelle 17).

Die Aussage steht im Widerspruch zu den Ergebnissen 1 und 2, die unter Einhaltung der Phasendifferenzierung generiert wurden. Die Phasendifferenzierung ist notwendig, damit Fehlinterpretationen vermieden werden und Effekte wie der Niveauverlust identifiziert werden können. Mit dem Wilcoxon Vorzeichen-Rangtest soll der Effekt des Niveauverlustes einer Signifikanzprüfung unterzogen werden (Tabelle 18).

Wilcoxon Vorzeichen-Rangtests auf Niveaustufenunterschiede des Experimentierens - Vergleich der Experimentierphasen 1 und 2					
Phase	Gruppe	N	N _{gültig}	Z	p-Wert
1 + 2	KG	24	18	3,72	< 0,001
1 + 2	TG 1	24	20	3,92	< 0,001
1 + 2	TG 2	10	10	2,80	< 0,001
1 + 2	TG 3	24	22	4,11	< 0,001
Legende: KG Kontrollgruppe TG Testgruppe: Reproduktion (TG 1), Reorganisation (TG 2), Transfer (TG 3) N Anzahl der Probanden der KG und TG N _{gültig} Anzahl der Probanden unter Eliminierung der Rangdifferenz Null (Nulldifferenz)					

Tab. 18: Wilcoxon Vorzeichen-Rangtests – mit gebundenen (betragsgleichen) Differenzen – auf Niveaustufenunterschiede des Experimentierens zwischen den Experimentierphasen 1 und 2 für die Test- und Kontrollgruppen. **Fettdruck:** signifikante Effekte für $p < 0,05$

Mit den Wilcoxon Vorzeichen-Rangtests wird geprüft, ob zwischen den Experimentierphasen 1 und 2 signifikante Niveaustufenunterschiede vorliegen. Die Ergebnisse sind für die Kontrollgruppe und für jede Testgruppe hochsignifikant (Tabelle 18), wodurch der identifizierte Niveaustufenabfall bzw. Niveauverlust (Niveaugap) statistisch gestützt wird (Ergebnis 3).

Das Berechnungsverfahren des Wilcoxon Vorzeichen-Rangtests liefert in Kombination mit den Ergebnissen der Tabelle 14 eine Erklärung, warum die Niveaustufenunterschiede zwischen den Experimentierphasen nicht nur signifikant, sondern hochsignifikant sind: Der Test analysiert Messdaten zweier Stichproben, die voneinander abhängig sind. Als ein Beispiel für den Einsatz des Wilcoxon Vorzeichen-Rangtests können Messwiederholungen genannt werden, bei denen die Messdaten von derselben Gruppe zu zwei unterschiedlichen Messzeitpunkten mit demselben Messinstrument erhoben werden (Rasch et al., 2006, S. 161 ff.). In Analogie zur Messwiederholung stehen die beiden Experimentierphasen in einem Abhängigkeitsverhältnis, denn der Zeitpunkt der Niveaustufenerhebung in Experimentierphase 1 überlagert sich nicht mit dem Zeitpunkt der Niveaustufenerhebung in Experimentierphase 2 und die Kriterien zur Ermittlung der Niveaustufen sind für beide Experimentierphasen identisch (vgl. Kap. 4.4.2). Im Gegensatz zur Messwiederholung muss der Vergleich der Niveaustufen aber als Messwiederholung mit Änderung der Experimentierbedingungen definiert werden, weil beim Übergang von der ersten zur zweiten Experimentierphase die Experimentiertätigkeit von der Arbeit mit vorgegebenem Experimentiermaterial zur Arbeit mit nicht vorgegebenem Experimentiermaterial (eigene Konstruktion) wechselt. Der Wilcoxon Vorzeichen-Rangtest misst die Unterschiede zwischen den Niveaustufen, die durch den Wechsel der Experimentierbedingungen verursacht werden. Diese Messung gibt den Niveauverlust (Niveaugap) zwischen den Experimentierphasen wieder.

Für alle Gruppen ist das Testergebnis hochsignifikant, weil der Wilcoxon Vorzeichen-Rangtest Differenzen von Messwertpaaren in eine Rangreihe bringt, die aus den Niveaustufendifferenzen zwischen der Experimentierphase 1 und 2 gebildet werden (Diehl & Arbinger, 2001, S. 565 ff). Zwischen den Niveaustufendifferenzen treten keine Vorzeichenwechsel auf, da kein Proband (Probandenpaar) in der Experimentierphase 2 eine höhere Niveaustufe als in der Experimentierphase 1 erreicht (Tabelle 14). Aus diesem Grund wird die Annahme der Nullhypothese, eine Gleichverteilung der positiven und negativen Rangdifferenzen verworfen. Folglich werden die Ergebnisse für alle Gruppen signifikant.

Kritische Bewertung der statistischen Auswertungsverfahren. Durch die quantitativen Berechnungen werden die qualitativen Ergebnisse der Niveaustufen des Experimentierens gestützt, statistisch jedoch nicht abgesichert. Die Daten sind kritisch zu bewerten, weil die statistischen Verfahren die Komplexität der Experi-

mentierprozesse nur unvollständig erfassen. Vor allem Zeiteffekte wie die Abfolge und die Dauer der Experimentierphasen werden von den Testverfahren nicht angemessen berücksichtigt. Sie können mit Hilfe der Niveaustufengrafiken (Abbildung 10) und den Experimentiertypenbeschreibungen (Kap. 4.3.3) nachgewiesen werden, wenn die Experimentiermuster der Typen A & B mit den Experimentiermustern der Typen C & D verglichen werden: Während bei den Typen A & B die Experimentierphase 2 immer nach Abschluss der Experimentierphase 1 einsetzt, sind bei Probandenpaaren, deren Experimentiertätigkeit nach den Mustern C & D verläuft, wiederholte Wechsel zwischen den Experimentierphasen und ein Einstieg in die Experimentiertätigkeit mit der eigenen Profilkonstruktion (Phase 2) nicht selten. Die Aussagekraft der Testverfahren könnte durch eine Kontrolle der Effekte erhöht werden, beispielsweise mit einer streng regulierten Experimentierabfolge und einer festen Vorgabe von Experimentierzeiten, so dass ein potentieller Einfluss der Zeiteffekte auf die Ergebnisse der Niveaustufen reduziert werden würde. Von den Maßnahmen wurde Abstand genommen, weil die Eingriffe in den Experimentierprozess zu einer Verschiebung des Forschungsinteresses und der Forschungsziele führen würden.

Die quantitativen, statistischen Daten sind aus einem weiteren Grund kritisch zu bewerten, weil die Testverfahren den wechselseitigen, paarweisen Einfluss der Probanden (Probandenpaare) beim Experimentieren nicht berücksichtigen. Um den Mangel zu kompensieren, müssten adäquate statistische Verfahren entwickelt und validiert werden, die für Rangdaten mit gebundenen Rängen auch die gegenseitige Abhängigkeit der Probanden mit aufnehmen. Die Entwicklung passgenauer Testverfahren könnte vermieden werden, wenn die Experimentierprozesse von Probanden (Einzelarbeit) und nicht von Probandenpaaren (Partnerarbeit) beobachtet werden würde, mit der Konsequenz, dass verbale Argumentationsstrukturen wie „Vergleich & Bewertung von Messfolgen“ oder „Strategien & Ideen zur Profilkonstruktion“ durch Interviewmethoden o. ä. erfragt werden müssten.

Im Vergleich zu der eingesetzten Methode der verbalen Selbstauskunft (Argumentationen) der Probandenpaare wird durch Interviewmethoden der Grad der Beobachterreaktivität zwischen Datenaufnehmer und Datengeber deutlich erhöht. Die Diskrepanz zwischen einer präziseren statistischen Auswertung und einer Minimierung der Beobachterreaktivität wurde zugunsten einer Reduktion von Beobachtereffekten bei der Datenaufnahme entschieden, so dass die Beobachtung von Probanden in Einzelarbeit im Rahmen der Studie ausgeschlossen wurde.

4.5 Korrelation und Triangulation der Ergebnisse: NAW-Test & Beobachtung

Der Vergleich der Experimentalgruppen hat gezeigt, dass die einmalige, einführende, informierende, explizite und vorbereitende Instruktion für den Umgang mit Variablen wirkt. Positive Effekte konnten für die Testgruppen der Teilstichprobe sowohl durch die Ergebnisse des NAW-Tests (Kap. 4.3) als auch durch die Ergebnisse der Beobachtung (Kap. 4.4) nachgewiesen werden. Weil die Wirkung der vorbereitenden Maßnahme mit zwei verschiedenen Messinstrumenten begutachtet wurde, ist eine Triangulation der Ergebnisse sinnvoll. Dazu werden die probandenpaarspezifischen Ergebnisse des NAW-Tests in Abhängigkeit der Messzeitpunkte präsentiert und mit den Niveaustufen des Experimentierens verglichen. Die Niveaustufen des Experimentierens geben die Ergebnisse der Beobachtung wieder. Um vorab die Ergebnisse auf Merkmalszusammenhänge zu prüfen, werden Korrelationsanalysen erstellt. Kritische Anmerkungen zur Korrelation und Triangulation der Ergebnisse werden im letzten Abschnitt gemacht. Für alle Abschnitte des Kapitels gilt, dass das Akronym NAW-Test stellvertretend für die Dimension Naturwissenschaftliche Arbeitsweisen des im Kapitel 4.2.2 vorgestellten Messinstruments Fach- & NAW-Test verwendet wird.

Korrelationsanalysen. Die aufgezeigten Korrelationsanalysen beschreiben den Zusammenhang zwischen den allgemeinen Fähigkeiten für den Umgang mit Variablen, die mit dem Testinstrument NAW-Test erfasst wurden und den kontextspezifischen Fähigkeiten für den Umgang mit Variablen beim Experimentieren, die mit dem Messinstrument Beobachtungsbogen erhoben wurden. Bei der Wahl des Zusammenhangmaßes muss berücksichtigt werden, dass die Daten des NAW-Tests Intervallskalengüte haben und die Niveaustufen des Experimentierens rangskalierte Daten sind. Ein geeignetes Maß für die Kombination der beiden Skalenniveaus ist der Rangkorrelationskoeffizient r_s nach Spearman (Rasch et al., 2006, S. 145). Für das intervallskalierte Merkmal „Allgemeine Fähigkeiten im Umgang mit Variablen“ hat die Korrelation mit einem rangskalierten Merkmal zur Konsequenz, dass es eine Herabstufung des Skalenniveaus erfährt. Für beide Rangreihen gilt, dass die Rangplätze von mehreren Probanden geteilt werden (Rangbindung bzw. verbundene Ränge). Die beiden Rangreihen unterscheiden sich durch die Anzahl der Ränge und durch das Verhältnis „Anzahl der verbundenen Ränge zur Anzahl aller Rangplätze“. Bei den Niveaustufen des Experimentierens verteilen sich die Rangplätze auf eine deutlich geringe Anzahl von Rängen als beim NAW-Test. Zudem variiert bei den Niveaustufen des Experimentierens die Anzahl der Ränge in Abhängigkeit von den Experimentierphasen. In der ersten Experimentierphase verteilen sich alle Rangplätze auf drei Ränge (Niveaustufen 1 – 3) und in

der zweiten Experimentierphase verteilen sich unter Vernachlässigung eines Rangplatzes alle Rangplätze auf zwei Ränge (Niveaustufen 1 – 2). Das hat zur Folge, dass in der Experimentierphase 1 nur verbundene Ränge und in der Experimentierphase 2 fast nur verbundene Ränge vorliegen. Dadurch überschreitet das Verhältnis „Anzahl der verbundenen Ränge zur Anzahl aller Rangplätze“ die kritische Grenze von 20 %, weshalb der Rangkorrelationskoeffizient r_s nicht mehr approximativ berechnet werden darf (Bortz, 2005, S. 233 f.).

Die Vorgabe ist für alle Rangkorrelationen bindend, die zwischen den Ergebnissen des NAW-Tests für die Messzeitpunkte 1 – 3 und den Ergebnissen des Experimentierens für die Experimentierphasen 1 und 2 erstellt werden (6 Rangkorrelationen). Von den Rangreihen des NAW-Tests wird die kritische Grenze ebenfalls überschreiten. Der Anteil an verbundenen Rängen zur Anzahl aller Rangplätze bewegt sich zwischen 50 % und 70 % und hängt vom Messzeitpunkt der Datenaufnahme ab (69 % zum Messzeitpunkt 1, 57 % zum Messzeitpunkt 2 und 53 % zum Messzeitpunkt 3).

Eine hohe Korrelation zwischen den Merkmalen ermöglicht eine wechselseitige Vorhersage der Ergebnisse. Eine Vorhersage von Ergebnissen wird akzeptiert, wenn ein zeitlich früher eintretendes Ereignis ein zeitlich nachfolgendes Ereignis beeinflussen kann. Beispielsweise gestattet eine hohe Korrelation zwischen den Ergebnissen des NAW-Tests zum Messzeitpunkt 1 und den Ergebnissen der Beobachtung eine Vorhersage der kontextspezifischen Fähigkeiten im Umgang mit Variablen beim Experimentieren (Kriteriumsvariable) durch die allgemeinen Fähigkeiten im Umgang mit Variablen (Prädiktorvariable), ohne dass die Probanden die offene Experimentieraufgabe bearbeitet haben müssen (Beispiel 1). In Umkehrung zum Beispiel 1 können die kontextspezifischen Fähigkeiten im Umgang mit Variablen beim Experimentieren, die mit dem Messinstrument Beobachtungsbogen erhoben wurden, nicht als Prädiktor für das Kriterium „Allgemeine Fähigkeiten im Umgang mit Variablen“ zum Messzeitpunkt 1 fungieren (Beispiel 2), weil zum Messzeitpunkt 1 das Messinstrument NAW-Test zeitlich vor dem Messinstrument Beobachtungsbogen eingesetzt wird (vgl. Kap. 4.2.1). Über die Beispiele 1 und 2 hinaus bietet das Design der Studie eine Hilfe, für welche Rangkorrelationen die Umkehrung von Prädiktor- und Kriteriumsvariablen akzeptiert, interpretiert und bewertet werden kann.

Die Bewertung der Korrelation wird durch den Grad des Merkmalszusammenhangs, der Signifikanz der Korrelation und der Effektstärke vorgenommen. Als Effektstärkenmaß kann beispielsweise der Korrelationskoeffizient benutzt werden. Zur Klassifikation der Effektstärke können die Grenzen der Produkt-Moment-Korrelation nach Pearson verwendet werden, weil der Rangkorrelationskoeffizient r_s nach Spearman mit dem Korrelationskoeffizienten r der Produkt-Moment-Korrelation identisch ist (Diehl & Arbinger, 2001, S. 619). Die Einordnung der

berechneten Korrelationskoeffizienten wird nach der Konvention von Cohen vorgenommen, wonach ein kleiner Effekt für $0,10 \leq r < 0,30$, ein mittlerer Effekt für $0,30 \leq r < 0,50$ und ein großer Effekt für $0,50 \leq r$ vorliegt (Rasch et al., 2006, S. 133).

Rangkorrelationskoeffizient r_s zwischen den Ergebnissen der NAW-Tests zu den Messzeitpunkten und den Niveaustufen des Experimentierens für die Experimentierphasen 1 und 2 der Teilstichprobe					
$r_s, N = 82$	Phase 1	Phase 2	$p, N = 82$	Phase 1	Phase 2
1	0,33	0,23	1	0,003	0,035
2	0,48	0,29	2	< 0,001	0,008
3	0,36	0,33	3	0,001	0,002
Legende: r_s Rangkorrelationskoeffizient nach Spearman N Anzahl der Probanden der Teilstichprobe 1 – 3 Messzeitpunkt: Vortest (1), Nachtest (2), Verzögerter Nachtest (3) Phase 1 Experimentierphase 1: Profilkonstruktion vorbereiten Phase 2 Experimentierphase 2: Profil konstruieren					

Tab. 19: Rangkorrelationskoeffizient r_s nach Spearman zwischen den Ergebnissen der NAW-Tests zu den Messzeitpunkten 1 – 3 und den Niveaustufen des Experimentierens für die Experimentierphasen 1 und 2 der Teilstichprobe. **Fettdruck:** signifikante Korrelationen für $p < 0,05$

Die Berechnung und Bewertung des Merkmalszusammenhangs (Tabelle 19) erfordert, dass vor den Korrelationsanalysen jedem Proband eines Probandenpaares das gemeinsam generierte Ergebnis des Experimentierens (Niveaustufen) zugewiesen wird. Die Zuweisung ist für die Datenauswertung notwendig, damit der Korrelationskoeffizient zwischen den allgemeinen Fähigkeiten im Umgang mit Variablen, die probandenweise erhoben wurden und den kontextspezifischen Fähigkeiten im Umgang mit Variablen beim Experimentieren, die probandenpaarweise erhoben wurden, nicht für die Kombination aus Probanden und Probandenpaaren bestimmt wird. Darüber hinaus ist die Zuweisung notwendig, weil die bekannten statistischen Auswertungsverfahren eine Berechnung des Korrelationskoeffizienten für Rangdaten mit verbundenen Rängen bei gegenseitiger Abhängigkeit der Probanden nicht unterstützen. Der Mangel an passgenauen statistischen Testinstrumenten könnte kompensiert werden, indem die Experimentierprozesse nur von Probanden erhoben werden, die nicht in Partnerarbeit experimentieren. Obwohl eine Änderung der Datenerhebungsmethode einen Vorteil

für die statistische Aufbereitung bietet, muss berücksichtigt werden, dass sie sich ergebnisverzerrend auf die Datengenerierung auswirken kann (vgl. Kap. 4.4.4, Kritische Bewertung der statistischen Auswertungsverfahren). Deshalb wurden die Experimentierprozesse von Probandenpaaren und nicht von einzelnen Probanden erhoben.

In der Tabelle 19 werden mit den sechs Rangkorrelationen die Merkmalszusammenhänge für die gesamte Teilstichprobe angegeben ($N = 82$ Probanden). Alternativ könnten die Korrelationsanalysen experimentalgruppenspezifisch (KG, TG 1 – 3) erstellt werden. Die Alternative wird aus zwei Gründen nicht angewandt. Erstens könnte eine experimentalgruppenspezifische Korrelationsanalyse zu der falschen Annahme führen, dass Probanden, die eine vorbereitende Instruktion erhalten haben, ein testspezifisches Training erhalten hätten. Zweitens ist mit einer experimentalgruppenspezifischen Korrelationsanalyse eine Selektion verbunden, die zur Konsequenz hat, dass bei der Erstellung des Merkmalszusammenhangs möglicherweise nicht die gesamte Variationsbreite der Teilstichprobe berücksichtigt wird.

Ergebnisse der Korrelationsanalysen. Die Tabelle 19 gibt die Rangkorrelationen des NAW-Tests zu den Messzeitpunkten 1 – 3 mit den Niveaustufen des Experimentierens für die Experimentierphasen 1 und 2 wieder. Für alle Korrelationsanalysen kann ein positiver Zusammenhang festgestellt werden. Die Rangkorrelationskoeffizienten liegen im Intervall $0 < r_s < 1$, was bedeutet, dass die beiden Rangreihen stochastisch weder unabhängig ($r_s = 0$) noch übereinstimmend ($r_s = 1$) sind. Ein tendenziell höherer Merkmalszusammenhang kann zwischen der Arbeit mit vorgegebenem Experimentiermaterial (Experimentierphase 1) und den Ergebnissen des NAW-Tests ermittelt werden als zwischen der Produktion eigener Profile (Experimentierphase 2) und den Ergebnissen des NAW-Tests. Der Korrelationskoeffizient erreicht ein Maximum für die Arbeit mit vorgegebenem Experimentiermaterial (Experimentierphase 1) und den Ergebnissen des NAW-Tests zum Messzeitpunkt 2. Dem Design (Kap. 4.2.1) kann entnommen werden, dass zu dem Messzeitpunkt alle Probanden die Bearbeitung der offenen Experimentieraufgabe abgeschlossen haben und je nach Gruppenzugehörigkeit (KG, TG) vor der offenen Experimentieraufgabe eine bzw. keine vorbereitende Instruktion zum Umgang mit Variablen erhalten haben.

Die Rangkorrelationskoeffizienten der Tabelle 19 werden durch die Ergebnisse ihrer Signifikanzprüfung erweitert. Mit der Nullhypothese eines Signifikanztests für Korrelationen wird geprüft, ob eine empirisch ermittelte Korrelation zweier Variablen aus einer Grundgesamtheit stammt, in der eine Korrelation von Null besteht. Die Alternativhypothese prüft das Gegenteil. Die Signifikanztests führen zu dem Ergebnis, dass alle berechneten Korrelationen auf dem 5 % bzw. 1 % Niveau bei ungerichteter Fragestellung statistisch abgesichert sind. Eine Klassifikation der

signifikanten Merkmalszusammenhänge führt nach der Konvention von Cohen zu einer mittleren Effektstärke für die Korrelationen zwischen den Ergebnissen der Experimentierphase 1 und den Ergebnissen des NAW-Tests und mehrheitlich zu kleinen Effektstärken für die Korrelationen zwischen den Ergebnissen der Experimentierphase 2 und den Ergebnissen des NAW-Tests.

Diskussion der Ergebnisse. An die Höhe des Korrelationszusammenhangs und in dessen Folge an die Grenzen des Effektstärkenmaßes könnte aufgrund des Feldcharakters der vorliegenden Studie ein geringerer Anspruch gestellt werden. Eine Abschwächung bzw. Minderung der Grenzen wird gegenüber Studien mit Laborcharakter nicht geltend gemacht, weil zur Kontrolle und Neutralisation potentieller Störvariablen zahlreiche Maßnahmen getroffen wurden (vgl. Kap. 4.2.1). Ferner hätten die Korrelationsanalysen zu erheblichen Abweichungen und in Folge dessen zu massiven Fehlinterpretationen geführt, wenn das Verhältnis der gebundenen Ränge zur Anzahl aller Rangplätze trotz Rangbindung approximativ berechnet worden wäre (Bortz, 2005, S. 233). Wie die folgenden Ergebnisse belegen, werden die Rangkorrelationskoeffizienten nach Spearman mit einer approximativen Berechnung für alle Messzeitpunkte (MZP) überschätzt (vgl. Tabelle 19):

$r_{S, N=82, \text{MZP } 1, \text{Phase } 1} = 0,43$; $r_{S, N=82, \text{MZP } 2, \text{Phase } 1} = 0,55$; $r_{S, N=82, \text{MZP } 3, \text{Phase } 1} = 0,46$;

$r_{S, N=82, \text{MZP } 1, \text{Phase } 2} = 0,33$; $r_{S, N=82, \text{MZP } 2, \text{Phase } 2} = 0,38$; $r_{S, N=82, \text{MZP } 3, \text{Phase } 2} = 0,42$.

Obwohl die Effektstärken zu den Korrelationen der Tabelle 19 nicht groß sind, ist die Interpretation erlaubt, dass sich die beiden Testverfahren stützen. Der NAW-Test und die Experimentierphase 1 stützen sich etwas stärker als der NAW-Test und die Experimentierphase 2, was nicht bedeutet, dass durch die Korrelationen eine wechselseitige Vorhersage der Ergebnisse gerechtfertigt wird. Das Fazit ist auf Basis von Literaturangaben zu bewerten. In Analogie zu der vorliegenden Studie kommen Shavelson, Ruiz-Primo & Wiley (1999), die zwischen paper-pencil-basierten und praktischen Experimentiertests einen Korrelationskoeffizienten von $r = 0,46$ (Schreiber et al., 2009) mit mittlerer Effektstärke bestimmt haben, zu dem Schluss, dass die beiden Testformate nicht austauschbar sind. Von mehreren Autoren ist zudem bekannt, dass die beiden Testformate nur schwach korrelieren (Ben-Zvi et al., 1977 nach Hammann et al., 2007, 2008; Comber & Keesee, 1973 nach Hammann et al., 2007, 2008; Robinson, 1969 nach Hammann et al., 2007, 2008; Tamir, 1972, 1974 nach Hammann et al., 2007, 2008). Die Tatsache, dass ein 30-jähriger Forschungsabschnitt nicht zu starken Merkmalsübereinstimmungen geführt hat, legitimiert die Frage, ob das Ziel, schriftliche und experimental-praktische Testformate austauschen zu wollen, richtig formuliert worden ist. Die Frage verliert ihre Berechtigung nicht, wenn Korrelationen mit höheren Effektstärken gefunden werden. Denn neben den übereinstimmenden Merkmalen der Testformate werden weiterhin nicht übereinstimmende Merkmale existent

bleiben (theoretische Ausnahme $r = 1$). Eine Alternative zur Substitution bietet die erkenntnisfördernde Triangulation von Testformaten, was im folgenden Abschnitt am Beispiel des schriftlichen Messinstruments NAW-Test und den Niveaustufen des Experimentierens aufgezeigt werden soll.

Triangulation. Das Ziel des Triangulationsberichts ist ein Erkenntnisgewinn, der durch den Vergleich von Ergebnissen entsteht, die mit verschiedenen Messinstrumenten ermittelt worden sind (NAW-Test & Beobachtung). Für diesen Zweck wird mit der Abbildung A 5a (Anhang) eine Präsentationsform gewählt, die sowohl eine experimentalgruppenspezifische als auch eine experimentalgruppenübergreifende Begutachtung korrespondierender Ergebnisse zulässt. Bei der Datenauswertung ist allerdings zu beachten, dass die Ergebnisse des NAW-Tests vom Messzeitpunkt abhängen. Während in der Abbildung A 5a nur die Ergebnisse des NAW-Tests zum Messzeitpunkt 1 mit den Niveaustufen des Experimentierens zusammengeführt werden, werden in den Abbildungen A 5b und A 5c (Anhang) die Ergebnisse des NAW-Tests zu den Messzeitpunkten 2 und 3 mit den Niveaustufen des Experimentierens trianguliert. Hierbei ist wichtig zu wissen, dass die nach den Ergebnissen des NAW-Tests zum Messzeitpunkt 1 getroffene Zuordnung der Probandenpaare in den oberen, mittleren und unteren Punktebereich für die Messzeitpunkte 2 und 3 nicht mehr angepasst bzw. aktualisiert werden. Die Maßnahme ermöglicht, dass bereichsgrenzenüberschreitende Abweichungen des NAW-Tests zwischen den Messzeitpunkten 2 und 1, den Messzeitpunkten 3 und 1 sowie den Messzeitpunkten 3 und 2 einfacher identifiziert werden können. Als ein Beispiel seien die Probandenpaare Nr. IX und Nr. XI der Kontrollgruppe genannt, die zum Messzeitpunkt 1 in den unteren Punktebereich eingewiesen wurden (vgl. Abbildung A 5a). Nach den Ergebnissen des NAW-Tests zum Messzeitpunkt 3 müsste das Probandenpaar Nr. IX in den oberen Punktebereich und der Proband P1 des Probandenpaares Nr. XI unterhalb des unteren Punktebereichs eingruppiert werden (vgl. Anhang, Abbildung A 5c). Die Änderungen im NAW-Test könnten jedoch weniger gut erkannt werden, wenn nach jeder Messwiederholung die Zuweisung der Punktebereiche erneuert werden würde. Zum orientierenden Vergleich dürfen die bereichsspezifischen Mittelwerte der Experimentalgruppen zum Messzeitpunkt 1 verwendet werden: 11,5 Punkte für den oberen Bereich, 9,5 Punkte für den mittleren Bereich und 7,5 Punkte für den unteren Bereich.

Der Triangulationsbericht wird auf Basis der Abbildungen des Anhangs (vgl. Abbildungen A 5a – c) chronologisch verfasst. Er beginnt mit der Darstellung der Erkenntnisse, die aus dem Vergleich des NAW-Tests zum Messzeitpunkt 1 mit den Niveaustufen des Experimentierens generiert werden können. Daran anschließend folgt ein Vergleich der Niveaustufen des Experimentierens mit den Ergebnissen des NAW-Tests zu den Messzeitpunkten 2 und 3. Der Schwerpunkt der Berichterstattung liegt auf dem Ergebnisvergleich zum Messzeitpunkt 1 und konzentriert

sich wegen der Unregelmäßigkeiten während der Datenerhebung in der Testgruppe 2 (Reproduktion) auf den Vergleich der beiden anderen Testgruppen mit der Kontrollgruppe.

Messzeitpunkt 1. Experimentierphase 1. In der Abbildung A 5a werden Probandenpaare der Experimentalgruppen verglichen, die aufgrund ihrer Ergebnisse im NAW-Vortest parallelisiert wurden (vgl. Kap. 4.3). Der Abbildung ist zu entnehmen, dass nur ein Drittel der Probandenpaare der Kontrollgruppe auf der höchsten Niveaustufe (Niveaustufe 3) und zwei Drittel der Probandenpaare auf einer tieferen Niveaustufe experimentieren (vgl. auch Kap. 4.4.4). Aus der Gegenüberstellung der Daten (Verteilung der Niveaustufen des Experimentierens über die Punktebereiche des NAW-Tests) werden für die Kontrollgruppe zwei Besonderheiten erkennbar. Erstens: Die vier Probandenpaare mit der höchsten Experimentierstufe befinden sich nicht nur im oberen Punktebereich, sondern verteilen sich über alle drei Punktebereiche des NAW-Tests. Zweitens: Ebenfalls überraschend ist die Verteilung der fünf Probandenpaare mit der Niveaustufe 2, die anders als erwartet, nicht nur im unteren und mittleren Punktebereich, sondern auch im oberen Punktebereich des NAW-Vortests vertreten sind. In Abgrenzung dazu sind die drei Probandenpaare mit der Experimentierniveaustufe 1 nur im mittleren und unteren Punktebereich zu finden. Für die Kontrollgruppe deutet diese Verteilung der Niveaustufen über die Punktebereiche des NAW-Vortests auf keine Regelmäßigkeit hin.

Den Probandenpaaren der Kontrollgruppe werden Probandenpaare der Testgruppen (TG 1, TG 3) gegenübergestellt, die im NAW-Vortest vergleichbare Resultate erzielt haben (Abbildung A 5a, Kap. 4.3). Nach den Ergebnissen der Abbildung A 5a und des Kapitels 4.4.4 experimentieren 10 von 12 Probandenpaaren der Testgruppe 1 und 9 von 12 Probandenpaaren der Testgruppe 3 in der Phase 1 auf der höchsten Niveaustufe (Niveaustufe 3). In beiden Testgruppen zusammen experimentieren nur fünf von 24 Probandenpaaren auf einer tieferen Niveaustufe. Die Verteilung der Niveaustufen über die Punktebereiche des NAW-Vortests ist in beiden Testgruppen ähnlich, weshalb eine gemeinsame Auswertung möglich ist. Es ist festzustellen, dass die Verteilung der Niveaustufen über den oberen und mittleren Punktebereich einer Regelmäßigkeit folgt. Für die Testgruppen ist charakteristisch, dass in dem oberen und mittleren Punktebereich nur Probandenpaare mit der höchsten Niveaustufe (Niveaustufe 3) zu finden sind. Von diesem Befund weicht ein Probandenpaar der Testgruppe 3 ab. Im unteren Punktebereich ist die Verteilung der Niveaustufen weniger regelmäßig, da dieser zur Hälfte von Probandenpaaren mit der Niveaustufe 3 und zur Hälfte von Probandenpaaren mit der Niveaustufe 2 bzw. 1 belegt wird.

Zusammenfassung und Diskussion der Ergebnisse (Messzeitpunkt 1, Experimentierphase 1). Die Ergebnisse des NAW-Tests liegen bei den Probanden der Teilstichprobe (N = 82) im Bereich von 5 – 13 Punkten (Wertebereich 0 – 13 Punkte). Trotz

der starken Schwankungen kann an den Testgruppen erkannt werden, dass ein Grundniveau an allgemeinen Fähigkeiten für den Umgang mit Variablen (NAW-Test) als Voraussetzung für einen niveauvollen Umgang mit Variablen beim Experimentieren (Niveaustufe 3) erreicht bzw. überschritten wird. Weiterhin zeigt ein Vergleich der Testgruppen (TG1, TG 3) mit der Kontrollgruppe, dass ein Fehlen der vorbereitenden Instruktion bei den meisten Probanden der Kontrollgruppe nicht zu einem niveauvollen Umgang mit Variablen beim Experimentieren führt (Niveaustufe 3), obwohl ein Grundniveau der allgemeinen Fähigkeiten für den Umgang mit Variablen erreicht bzw. überschritten wird. Diese Erkenntnis (Erkenntnis 1) gibt einen Hinweis darauf, dass ein Wirkungszusammenhang zwischen den allgemeinen Fähigkeiten, der vorbereitenden Instruktion und den kontextspezifischen Fähigkeiten beim Experimentieren (Niveaustufen des Experimentierens) bestehen muss. Im Kontrast zur Erkenntnis 1 stehen die Ergebnisse von fünf Probandenpaaren der Testgruppen 1 und 3, die mehrheitlich in den unteren Punktebereich des NAW-Tests eingruppiert wurden und beim Experimentieren nicht die höchste Niveaustufen erreicht haben (vgl. Abbildung A 5a, Probandenpaare Nr. X und XI der TG 1 sowie die Probandenpaare Nr. VIII, X und XII der TG 3). Die Wirksamkeit der vorbereitenden Instruktion wird dadurch nicht bezweifelt, weil es Probanden bzw. Probandenpaaren mit vergleichbaren Ergebnissen im NAW-Test gelungen ist, beim Experimentieren die höchste Niveaustufe zu erreichen (vgl. Abbildung A 5a, Probandenpaare Nr. IX und XII der TG 1 sowie Probandenpaare Nr. IX und XI der TG 3). Als eine mögliche Erklärung wird vermutet, dass von den Probanden Experimentierstrategien wie die Variablenkontrollstrategie noch nicht hinreichend beherrscht werden. Eine alternative Erklärung könnte sein, dass die Probanden die Variablenkontrollstrategie und die Aufnahme bzw. Auswertung nicht-konfundierender Messfolgen durchaus verstehen, aber den Einsatz dieser Strategie bei der Bearbeitung der offenen Experimentieraufgabe für nicht notwendig oder zielführend halten. Diese Erklärung ist nicht unbegründet, da van Joolingen & de Jong (1997) eine analoge Erklärung für den Umgang mit Hypothesen geben, nach der Probanden das machen, was sie aus der eigenen Ansicht heraus für zielführend halten.

Messzeitpunkt 1. Experimentierphase 2. Der für die Experimentierphase 1 mit der Erkenntnis 1 formulierte Hinweis auf den Zusammenhang zwischen den allgemeinen Fähigkeiten für den Umgang mit Variablen, der vorbereitenden Instruktion und den Ergebnissen des Experimentierens (Niveaustufen) kann für die Experimentierphase 2 nicht hergestellt werden. Als Ursache kann der zwischen den beiden Experimentierphasen festgestellte Niveauverlust (Niveaugap, vgl. Kap. 4.4.4) benannt werden, der für den Vergleich der Ergebnisse des NAW-Tests mit den Niveaustufen des Experimentierens eine differenzierte Auswertung erfordert. Der Abbildung A 5a ist zu entnehmen, dass die Niveaustufe 2 bei der Testgruppe 1 im

oberen und mittleren Punktebereich, bei der Testgruppe 3 im oberen Punktebereich und bei der Kontrollgruppe in keinem Punktebereich des NAW-Tests dominiert. Demgegenüber ist die Niveaustufe 1 bei der Testgruppe 1 in keinem Punktebereich, bei der Testgruppe 3 im unteren Punktebereich und bei der Kontrollgruppe in allen Punktebereichen des NAW-Vortests dominant. Hierzu muss ergänzt werden, dass im unteren Punktebereich der Testgruppe 1 und im mittleren Punktebereich der Testgruppe 3 die Niveaustufen 1 und 2 je zur Hälfte vertreten sind.

Zusammenfassung und Diskussion der Ergebnisse (Messzeitpunkt 1, Experimentierphase 2). Der hohe Anteil an Probanden bzw. Probandenpaaren, die in der Kontrollgruppe die Niveaustufe 1 beim Experimentieren erreicht haben, ist besonders überraschend, denn: Als ein Ergebnis der Münchener Längsschnittstudie LOGIK wird berichtet, dass über 70 % der 17-Jährigen Versuchspersonen die Variablenkontrollstrategie spontan verwenden (Bullock & Sodian, 2003). Anders als in der LOGIK-Studie gelingt das über 80 % der Probandenpaare (mit einem Durchschnittsalter von 16 Jahren und 2 Monaten, vgl. Kap. 4.2.1) nicht, wenn sie ohne vorbereitende Instruktion experimentieren (10 von 12 Probandenpaare der Kontrollgruppe erzielen in der Experimentierphase 2 die Niveaustufe 1, vgl. auch Kap. 4.4.4). Als eine mögliche Ursache der kontrastiven Ergebnisse werden Unterschiede zwischen den eingesetzten Experimentierumgebungen vermutet. Im Gegensatz zur LOGIK-Studie mussten die Probanden in dieser Studie den Umgang mit Variablen in Realexperimenten zeigen.

Messzeitpunkt 2. Experimentierphasen 1 und 2. Eine Erhebung über die Änderungen der kontextspezifischen Fähigkeiten im Umgang mit Variablen beim Experimentieren liegt außerhalb des Untersuchungsinteresses, weshalb das Messinstrument Beobachtungsbogen im Verlauf der Studie nicht wiederholt eingesetzt wird. Für die Triangulation der Ergebnisse hat das zur Folge, dass die Niveaustufen des Experimentierens von der Abbildung A 5a auf die Abbildungen A 5b und A 5c (Anhang) übertragen werden müssen, damit sie mit den korrespondierenden Ergebnissen des NAW-Tests zu den Messzeitpunkten verglichen werden können. Da die Unterschiede zwischen den Messzeitpunkten der Experimentalgruppen mit Hilfe der t-Tests statistisch bereits hinreichend dokumentiert werden (Kap. 4.3), ist es das Ziel dieses Abschnitts, die Unterschiede durch die Einteilung der Probanden bzw. Probandenpaare in den oberen, mittleren und unteren Punktebereich zu differenzieren.

Die Differenzierung führt zu der Erkenntnis (Erkenntnis 2), dass die vorbereitende Instruktion zu einer Stabilisierung und Steigerung der Ergebnisse im NAW-Test führt. Die stabilisierende bzw. ergebnissteigernde Wirkung ist nicht auf Subgruppen wie Probanden des oberen, mittleren und unteren Punktebereichs beschränkt. Zuwächse können sogar bei den wenigen Probanden der Testgruppen (TG 1, TG 3) identifiziert werden, die in der Experimentierphase 1 nicht die höchste Niveaustufe

erreicht haben (Probanden der Probandenpaare Nr. X und XI der TG 1 sowie Probanden der Probandenpaare Nr. VIII, X und XII der TG 3). Insgesamt erzielten 50 % der Probanden in der Testgruppe 1 und über 60 % der Probanden in der Testgruppe 3 eine Ergebnissteigerung. Für zwei der 24 Probanden der Testgruppe 1 und für einen der 24 Probanden der Testgruppe 3 verschlechtert sich das Ergebnis des NAW-Tests mit der Messwiederholung.

In der Kontrollgruppe liegt ein anderes Verhältnis vor: Ein Drittel der Probanden (8) erreicht im Nachtest eine höhere Punktzahl als im NAW-Vortest, etwa ein Drittel der Probanden (7) erreicht im Nachtest eine kleinere Punktzahl als im NAW-Vortest und etwa ein Drittel der Probanden (9) erreicht zu beiden Messzeitpunkten dieselbe Punktzahl. Die Fluktuationen, die mit dem Wechsel der Messzeitpunkte einhergehen, sind nicht auf Subgruppen wie Probanden des oberen, mittleren oder unteren Punktebereichs oder auf Probanden der systematischen, semi-systematischen oder unsystematischen Niveaustufe des Experimentierens beschränkt. Dennoch ist eine Regelmäßigkeit für die Probanden der Kontrollgruppe erkennbar, deren Punktzahlen vom NAW-Vortest zum NAW-Nachtest abnehmen: Sie erreichen in der Experimentierphase 1 nicht die höchste Niveaustufe, sondern nur die mittlere oder untere Niveaustufe des Experimentierens. Das trifft für sechs der sieben Probanden zu (Proband P1 der Probandenpaare Nr. II, VI, X und XI sowie Proband P2 der Probandenpaare Nr. II und X; vgl. Abbildungen A 5a und A 5b). Zu den genannten Probanden der Kontrollgruppe muss ergänzt werden, dass die Punktverluste moderat ausfallen und nicht um mehr als zwei Maßeinheiten zum NAW-Vortest differieren. Aber nicht nur für die Probanden der Kontrollgruppe, sondern auch für die Probanden der Testgruppen (TG 1, TG 3) sind die meisten Abweichungen im NAW-Test moderat und nehmen in Abhängigkeit des Messzeitraums höchstens um zwei Punkte zu bzw. ab.

Messzeitpunkt 3. Experimentierphasen 1 und 2. Wenn die Ergebnisse des NAW-Tests zum Messzeitpunkt 3 mit den Ergebnissen des NAW-Tests zum Messzeitpunkt 2 verglichen werden, dann kann in den Testgruppen 1 und 3 eine steigende Anzahl an Probanden mit moderaten Punktverlusten wahrgenommen werden. Die Punktverluste sind bei Probanden zu registrieren, die dem oberen, mittleren und unteren Punktebereich zugeordnet wurden und in der ersten Experimentierphase die höchste Niveaustufe erzielt haben. Die Ergebnisse der Experimentalgruppen werden durch die Datenauswertung des Kapitels 4.3 hinreichend beschrieben. Dazu zählen auch die Unregelmäßigkeiten, die in der Testgruppe 2 vorliegen (80 % der Probanden steigern vom Nach- zum Verzögerten Nachtest ihre Ergebnisse).

Kritische Anmerkungen zur Korrelation und Triangulation. Die Korrelationsanalysen geben keine Auskunft über die Kausalitätsbeziehungen der untersuchten Merkmale. Der Kausalzusammenhang zwischen den Ergebnissen des NAW-Tests,

der vorbereitenden Instruktion und den Niveaustufen des Experimentierens wird im Abschnitt Triangulation aufgenommen und fließt in die beiden Erkenntnisse des Triangulationsberichts mit ein. Um den kausalen Zusammenhang zwischen den drei Merkmalen statistisch zu belegen, müssten Partialkorrelationsanalysen erstellt werden, mit dem Ziel, die vorbereitende Instruktion als einflussreiche Drittvariable zu partialisieren. Da vermutet werden kann, dass die vorbereitende Instruktion mit den Niveaustufen des Experimentierens stärker korreliert als mit den Ergebnissen des NAW-Tests, würde die vorbereitende Instruktion die Funktion einer Suppressorvariablen einnehmen. Dies hätte zur Konsequenz, dass die resultierenden Partialkorrelationen gegenüber den Korrelationen der Tabelle 19 erhöht werden würden (Rasch, et al., 2006, S. 138). Eine derartige statistische Auswertung macht aber die Erfassung der Drittvariablen mit verschiedenen Ausprägungen in allen Beobachtungseinheiten notwendig, was über die Ziele der vorliegenden Studie hinaus gehen würde.

5 Hypothesen

Aus den Ergebnissen dieser Studie werden folgende Hypothesen generiert:

1. Die Fähigkeit zur naturwissenschaftlichen Arbeitsweise des Umgangs mit Variablen erlernen Schüler nicht ausschließlich durch die Tätigkeit des Experimentierens. Hierzu benötigen sie geeignete Unterstützungsmaßnahmen wie die einmalige, einführende, informierende, vorbereitende und explizite Instruktion. Dies trifft auch für den Fall zu, wenn die Schüler mit vorgefertigtem Experimentiermaterial arbeiten.
2. Die Unterstützung in Form einer einmaligen, einführenden, informierenden, vorbereitenden und expliziten Instruktion ist für die Arbeit mit vorgefertigtem Experimentiermaterial auch unter der Bedingung zielführend, dass die Schüler einen Kontexttransfer von der vorbereitenden Instruktion auf die (offene) Experimentieraufgabe leisten müssen.
3. Veränderte Experimentierbedingungen (Arbeit mit vorgefertigtem Experimentiermaterial/eigene Profilkonstruktion) führen auch ohne Kontexttransfer zu einem Niveauverlust (Niveaugap) beim Experimentieren. Dieser verstärkt sich, wenn zusätzlich von der vorbereitenden Instruktion auf die (offene) Experimentieraufgabe ein Kontexttransfer geleistet werden muss.
4. Die Offenheit einer Experimentieraufgabe hat einen positiven Einfluss auf den Lernerfolg, sofern Offenheit nicht in allen Dimensionen (Öffnungsmöglichkeiten) gleichzeitig besteht und eine Anhäufung zu vieler Ziele innerhalb der Aufgabenstellung vermieden wird.

6 Zusammenfassung und Ausblick

Das Ziel dieser Forschungsarbeit ist die Bewertung von Experimentierprozessen, die Gymnasiasten am Ende der SI (G 9) hinsichtlich der naturwissenschaftlichen Arbeitsweise Umgang mit Variablen bei der Bearbeitung einer offenen Experimentieraufgabe in Physik zeigen. Bewertet werden die beobachteten Prozesse durch eine kriteriengeleitete Aufstellung von Niveaustufen des Experimentierens. Die vergleichende Studie möchte durch die Zuordnung von Probanden in diese Niveaustufen die Wirksamkeit einer offenen Experimentieraufgabe aufzeigen und zwar unter der Bedingung, dass vor dem Experimentieren eine bzw. keine vorbereitende Schulung (Instruktion) für den Umgang mit Variablen durchgeführt wurde (Forschungsfrage 1).

Ein weiteres Untersuchungsinteresse liegt darin, zu erfassen, inwieweit die Probanden in der Lage sind, einen Kontexttransfer von der einmaligen, einführenden, informierenden, vorbereitenden und expliziten Instruktion auf die offene Experimentieraufgabe zu leisten (Forschungsfrage 2). Um die Graduierung der Kontextunterschiede anzuzeigen, werden die Begriffe Reproduktion (gleiche Kontexte), Reorganisation (ähnliche Kontexte) und Transfer (verschiedene Kontexte) verwendet. Zur Erzeugung der Kontextunterschiede werden die Ausprägungen des vermittelten Variablenkonzeptes für die Testgruppen TG 1 (Reproduktion), TG 2 (Reorganisation) und TG 3 (Transfer) variiert, wohingegen von allen Experimentalgruppen dieselbe offene Experimentieraufgabe zu bearbeiten ist. Um bei variierenden Kontexten die Vergleichbarkeit einer vorbereitenden Instruktion zu garantieren, wird ein selbst erstelltes Vermittlungskonzept für den Umgang mit Variablen benutzt (vgl. Kap. 3.2), dessen Inhalte, Struktur und Phasenablauf auf alle Konzeptausprägungen übertragen werden. Die offene Experimentieraufgabe verlangt von den Probanden eine Profilkonstruktion für das Modell eines Windrades, so dass bei den gegebenen Versuchsbedingungen eine möglichst hohe Spannung erzeugt wird. Dieses Aufgabenbeispiel wurde nach dem eigenen Ansatz des offenen Experimentierens entwickelt (vgl. Kap. 2.1 & 2.3).

Als ein zentrales Ergebnis kann festgehalten werden, dass die Fähigkeit der Probanden zum Transfer des Variablenkonzeptes nicht vom Grad der Kontext-

unterschiede zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe abhängt, da fast alle Probanden der Testgruppen die höchste Niveaustufe beim Experimentieren erreichen. Mit Ausnahme der Testgruppe 2, für die nur 10 Probanden rekrutiert werden konnten, standen in den beiden anderen Testgruppen jeweils 24 Probanden zur Verfügung. Davon konnten je nach Gruppenzugehörigkeit 18 – 20 Probanden in die höchste, systematische Niveaustufe (Niveaustufe 3), zwei bis vier Probanden in die mittlere, semi-systematische Niveaustufe (Niveaustufe 2) und zwei bis vier Probanden in die unterste, unsystematische Niveaustufe (Niveaustufe 1) gruppiert werden. Diese qualitativen Ergebnisse werden durch statistische Analysen gestützt, die auf der Basis von Mann-Whitney U-Tests für unabhängige Stichproben mit Rangbindung zu nicht signifikanten Unterschieden zwischen den Testgruppen führen. Demgegenüber werden die U-Tests statistisch signifikant, wenn die Kontrollgruppe (KG) paarweise mit den Testgruppen verglichen wird. Die statistischen Ergebnisse erlauben die Interpretation, dass es Schülern am Ende der S I ohne vorbereitende Instruktion (KG) nicht gelingt, effektiv beim Experimentieren mit Variablen umzugehen. Dieses Fazit geht mit der Verteilung der 24 Probanden der Kontrollgruppe über die Niveaustufen des Experimentierens einher, von denen nur 8 Probanden die Niveaustufe 3, 10 Probanden die Niveaustufe 2 und sechs Probanden die Niveaustufe 1 erreichen. Die Besetzung der Niveaustufen besagt, dass die Kriterien der höchsten Niveaustufe, wie die Aufnahme und Auswertung nicht-konfundierender Messfolgen, von den meisten Probanden der Kontrollgruppe beim Experimentieren nicht erfüllt werden können.

Die referierten Ergebnisse müssen mit Änderung der Experimentierbedingungen, unter denen die Experimentierprozesse bewertet werden, revidiert werden. Während die vorangehenden Resultate die Gruppenunterschiede für den Umgang mit Variablen beim Experimentieren beschreiben, wenn die Probanden vorgefertigte Versuchsmaterialien benutzten (Experimentierphase 1), geben die folgenden Ergebnisse die Gruppenunterschiede beim Experimentieren für den Umgang mit Variablen wieder, wenn die Probanden eigene Produkte herstellen (Experimentierphase 2). Hierbei ist es wichtig zu wissen, dass den Probanden der Testgruppen im Vorfeld des Experimentierens die naturwissenschaftliche Arbeitsweise nur am Beispiel vorgefertigter Versuchsmaterialien und nicht für die Herstellung eigener Produkte vermittelt wurde. Unter diesen Voraussetzungen liegt das Untersuchungsinteresse der zweiten Forschungsfrage auch darin, zu erfassen, inwieweit die Probanden die Fähigkeit zum Transfer des Variablenkonzeptes von einer vorbereitenden Instruktion auf eine offene Experimentieraufgabe haben, wenn sie zusätzlich zum Kontexttransfer einen weiteren Transfer, der durch die Änderung der Experimentierbedingungen hervorgerufen wird, leisten müssen.

Als ein zentrales Ergebnis kann festgehalten werden, dass die Fähigkeit zum Transfer des Variablenkonzeptes vom Grad der Kontextunterschiede zwischen der

vorbereitenden Instruktion und der offenen Experimentieraufgabe abhängt, wenn die Experimentierenden der Testgruppen neben veränderten Kontexten auch veränderte Experimentierumgebungen („Herstellung eigener Produkte“) vorfinden. Da diese Erkenntnis an den Besetzungsverhältnissen zwischen der semi-systematischen und der unsystematischen Niveaustufe der Experimentalgruppen mit gleicher Probandenzahl (TG 1, TG 3, KG) gut erkannt werden kann, wird die Niveaustufenverteilung der Testgruppe 2 nicht aufgezeigt. Von den 24 Probanden der Testgruppe 1 haben 18 die semi-systematische Experimentierstufe (Niveaustufe 2) erreicht. Das sind fast doppelt so viele Probanden wie in der Testgruppe 3 (10 Probanden) und mehr als viermal so viele Probanden wie in der Kontrollgruppe (vier Probanden). Demgegenüber wird die unsystematische Experimentierstufe (Niveaustufe 1) von vier Probanden der Testgruppe 1 (Reproduktion), von 14 Probanden der Testgruppe 3 (Transfer) und von 20 Probanden der Kontrollgruppe (ohne Vorbereitung) erreicht. Die systematische Experimentierstufe (Niveaustufe 3) kann unter den 82 Probanden der Teilstichprobe, deren Experimentierprozesse in beiden Experimentierphasen analysiert und bewertet wurden, für die Experimentierphase 2 insgesamt nur zweimal vergeben werden. Die beiden Probanden gehören der Testgruppe 1 an. Diese qualitativen Ergebnisse werden durch U-Tests statistisch gestützt, die nicht immer zu signifikanten Ergebnissen führen. Signifikant werden die Gruppenunterschiede zwischen der Kontrollgruppe und der Testgruppe 1, zwischen der Kontrollgruppe und der Testgruppe 2 sowie zwischen den Testgruppen 1 und 3. Nicht signifikant sind die Unterschiede zwischen der Kontrollgruppe und der Testgruppe 3, zwischen den Testgruppen 1 und 2 sowie zwischen den Testgruppen 2 und 3. Die signifikanten Ergebnisse erlauben die Interpretation, dass die Probanden die Fähigkeit zum Transfer des Variablenkonzeptes von der vorbereitenden Instruktion mit vorgefertigten Versuchsmaterialien auf die Herstellung eigener Produkte beim offenen Experimentieren haben. Aus der Kombination mit den nicht signifikanten Ergebnissen kann allerdings erkannt werden, dass die Fähigkeit der Probanden zum Transfer des Variablenkonzeptes vom Grad der Kontextunterschiede abhängt, wenn die Experimentierenden eigene Produkte herstellen sollen (Experimentierphase 2). Beispielsweise besagt der nicht signifikante Unterschied zwischen der Kontrollgruppe und der Testgruppe 3, dass denjenigen Probanden, die zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe verschiedene Kontexte vorfinden, der Transfer des Variablenkonzeptes von der vorbereitenden Instruktion mit vorgegebenem Versuchsmaterial auf die eigene Herstellung eines Produktes beim offenen Experimentieren nicht besser gelingt, als den Probanden der Kontrollgruppe, die keine vorbereitende Instruktion erhalten haben.

Die beiden vorausgehenden zentralen Ergebnisse basieren auf den Verteilungen der Niveaustufen unter Auflösung von zwei Experimentierphasen. Für eine angemessene Bewertung der Experimentierprozesse sind ebenfalls die Gruppen-

unterschiede von Bedeutung, die sich aus der Kombination der Niveaustufen durch Vergleich der ersten und zweiten Experimentierphase ergeben. Als ein zentrales Ergebnis kann herausgestellt werden, dass für den Wechsel von der ersten zur zweiten Experimentierphase bei allen Experimentalgruppen (KG, TG 1 – 3) ein statistisch hochsignifikanter Niveauverlust (Synonyme: Niveaugap, Niveauabfall) gefunden werden kann, falls die Niveaustufenunterschiede mit dem Wilcoxon Vorzeichen-Rangtest geprüft werden. Als Ursache für die abnehmende Fähigkeit zum Transfer des Variablenkonzeptes kann die Änderung der Experimentierbedingung benannt werden, die durch den Wechsel von der Arbeit mit vorgegebenem Versuchsmaterial (Experimentierphase 1) zur Herstellung eigener Produkte (Experimentierphase 2) geschaffen wird. Im Gegensatz zu den quantitativ-statistischen Ergebnissen kann aus den qualitativen Daten ausgelesen werden, dass die Stärke des Niveauverlustes mit dem Grad der Kontextunterschiede zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe variiert. Die Probanden der Testgruppe 3 (Transfer) zeigen einen größeren Niveauabfall als die Probanden der Testgruppe 1 (Reproduktion), der insgesamt für die Probanden, die keine vorbereitende Instruktion erhalten haben (KG), am stärksten ausgeprägt ist. Eine Folge nach steigendem Niveauverlust kann beispielsweise an der Verteilung der Niveaustufenkombinationen 3 und 2 sowie 2 und 1 (Niveaustufe in Phase 1, Niveaustufe in Phase 2) erkannt werden, obwohl die Niveaudifferenzen zwischen beiden Kombinationen identisch sind (Synonyme Schreibweise: 3/2; 2/1). Hierbei ist aber zu beachten, dass die Niveaustufenkombination 3/2 sowohl das höhere Ausgangsniveau (Phase 1) als auch das höhere Endniveau (Phase 2) hat. Die Niveaustufenkombination 3/2 wird in der Testgruppe 1 von 18 Probanden, in der TG 3 von 10 Probanden und in der Kontrollgruppe von vier Probanden erreicht. Eine reversible Tendenz, die der Verteilung der Niveaustufenkombination 3/2 nicht widerspricht, ergibt sich aus der Verteilung der Niveaustufenkombinationen 2/1, die in der Testgruppe 1 von zwei Probanden, in der Testgruppe 3 von vier Probanden und in der Kontrollgruppe von 10 Probanden belegt werden. Die übrigen, an dieser Stelle nicht aufgeführten Niveaustufenkombinationen (3/3, 3/1, 1/1), verstärken die Ausprägung der erstellten Folge nach steigendem Niveauverlust (TG 1, TG 2, TG 3, KG).

In einem Zwischenfazit kann festgehalten werden, dass es im Rahmen dieser Forschungsarbeit gelungen ist, einen aus der Literatur bisher nicht bekannten Niveauverlust beim Experimentieren zu identifizieren, der bei den Probanden während der Bearbeitung einer offenen Experimentieraufgabe durch einen Wechsel der Experimentierbedingungen ausgelöst worden ist. Die Interpretation des gefundenen Effektes ist eine zentrale Voraussetzung, um die Niveauunterschiede zwischen den Experimentierphasen durch weitere Unterstützungsmaßnahmen zu kompensieren. Die Wirkung der einmaligen, einführenden, informierenden, vorbereitenden und expliziten Instruktion wird hierdurch nicht in Frage gestellt. Damit

aber Fehlinterpretationen vermieden werden können, müssen u. a. die Bewertungskriterien und -verfahren berücksichtigt werden, die für die Einordnung in die Niveaustufen ausschlaggebend waren. Wesentlich ist hierbei, dass für die Bewertung der Experimentierprozesse eine Gesamtniveaustufe vergeben wurde, die aus den Prozessen der Datenaufnahme und aus den Prozessen der Datenauswertung ermittelt und für jede Experimentierphase separat vergeben wurde. Immer dann, wenn zwischen der Datenaufnahme und der Datenauswertung Niveauunterschiede vorlagen, richtete sich die Gesamtniveaustufe nach der tieferen der beiden Einzelniveaustufen (nivellierender Effekt). Es fiel jedoch auf, dass niemals die Prozesse der Datenaufnahme für den nivellierenden Effekt ausschlaggebend waren. Diese Erkenntnis wird von dem begleitenden Experimentiermuster unterstützt (vgl. Experimentiertyp A). Als ein Beispiel, das vor allem in der Experimentierphase 2 zu beobachten war und bei der Vergabe der Niveaustufenkombination 3/2 wirksam wurde, sei die messdatenbasierte Herstellung eigener Produkte und die Aufnahme nicht-konfundierender Messfolgen (Kriterium Niveaustufe 3, Datenaufnahme) mit „paarweisen Vergleichen“ der Messergebnisse (Kriterium Niveaustufe 2, Datenauswertung) genannt. Die Bewertungskriterien lassen erkennen, dass der Niveauabfall zwischen den Experimentierphasen keiner einheitlichen Interpretation bzw. Erklärung folgt. Die folgenden Interpretationen konzentrieren sich vor allem auf die Probanden der Testgruppen (TG 1 – 3), die nach einer vorbereitenden Instruktion die Experimentierstufenkombinationen 3/2 und 3/1 erreicht haben. In allen drei Testgruppen zusammen handelt es sich hierbei um etwas mehr als 75 % der Probanden, deren gruppenspezifische Verteilung von den Kontextunterschieden zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe beeinflusst wird.

Annähernd bei der Hälfte der Probanden mit der Niveaustufenkombination 3/2 (Verteilung s. o.) konnte in der zweiten Experimentierphase der nivellierende Effekt festgestellt werden, der aufgrund von Defiziten der Probanden bei der Auswertung ihrer Messergebnisse die Gesamtniveaustufe von der systematischen auf die semi-systematische Experimentierstufe sinken ließ. Das bedeutet, dass die Probanden bei der Herstellung eigener Produkte nicht-konfundierende Messfolgen aufgenommen haben, wobei die Variablenkontrollstrategie eingehalten wurde. Mit Hilfe weiterer Unterstützungsmaßnahmen könnte vermutlich das Experimentierverhalten der Probanden von der semi-systematischen auf das systematische Experimentierniveau gehoben werden. In Abgrenzung dazu konnten etwa bei der anderen Hälfte der Probanden mit der Niveaustufenkombination 3/2 keine Niveauunterschiede zwischen der Datenaufnahme und der Datenauswertung gefunden werden, was besagt, dass die Probanden bei der Herstellung eigener Produkte nicht gegen die Variablenkontrollstrategie verstoßen haben. Allerdings waren die Ausprägungen der generierten Messfolgen (Anzahl der Messwerte) nicht ausreichend genug, um den

Einfluss und den Verlauf (linear, extremal, ohne Einfluss) der untersuchten Variablen differenziert zu bewerten, weshalb die Prozesse der Datenaufnahme (und der Datenauswertung) nur dem semi-systematischen Niveau entsprachen. Da die aufgenommenen Messdaten einen umfassenden Vergleich von mehreren Messfolgen nicht unterstützten, bleibt ungewiss, ob auch bei diesen Probanden Defizite im Bereich der Datenauswertung vorliegen.

Die vorausgehenden Befunde können für die Erklärung des Niveauabfalls bei der Niveaustufenkombination 3/1 nicht vereinnahmt werden, weil die Experimentierprozesse zwischen den beiden Phasen wesentlich kontrastiver verlaufen, als bei der Niveaustufenkombination 3/2. Kennzeichnend für die Niveaustufenkombination 3/1 ist die unsystematische Arbeitsweise in der Experimentierphase 2, die nach den Kriterien zur Vergabe der Niveaustufe durch die Aufnahme und Auswertung von konfundierenden Messfolgen, Ansätzen konfundierender Messfolgen oder Einzelmessungen geprägt ist. Weiterhin wird aus dem begleitenden Experimentiermuster (vgl. Experimentiertyp B) ersichtlich, dass die unsystematische Arbeitsweise temporär von einer semi-systematischen Arbeitsweise abgelöst wird. Diese Instabilität des Experimentierverhaltens erlaubt die Interpretation, dass die Probanden die Aufnahme von nicht-konfundierenden Messfolgen entweder für nicht zielführend halten oder noch unentschieden sind, welche Strategien zielführend sind, wenn sie eigene Produkte herstellen. Ein Missverständnis läge vor, wenn den Probanden die Fähigkeit für den Umgang mit Variablen abgesprochen werden würde, weil sie für die Arbeit mit vorgegebenem Versuchsmaterial gezeigt haben, dass sie die Strategien zur Aufnahme und Auswertung nicht-konfundierender Messdaten beherrschen (Niveaustufe 3 in Experimentierphase 1). Es sei darauf verwiesen, dass van Joolingen & de Jong (1997) eine analoge Erklärung für den Umgang mit Hypothesen geben, nach der Probanden das machen, was sie aus der eigenen Ansicht heraus für zielführend halten. Eine andere Erklärung, mit der die Verteilung der Niveaustufenkombinationen in den Testgruppen stärker berücksichtigt werden würde, ist, dass den Probanden der Transfer des Variablenkonzeptes weniger gut gelingt, wenn der Grad der Kontextunterschiede zwischen der vorbereitenden Instruktion und der offenen Experimentieraufgabe zunimmt, denn: die Niveaustufenkombination 3/1 wurde unter den 24 Probanden der Testgruppe Reproduktion (TG 1) keinmal, unter den 10 Probanden der Testgruppe Reorganisation (TG 2) zweimal und unter den 24 Probanden der Testgruppe Transfer (TG 3) achtmal vergeben. Ergänzend sei hinzugefügt, dass diese Kombination unter den 24 Probanden der Kontrollgruppe viermal vergeben wurde. Vor allem die hohe Präsenz der Niveaustufenkombination in der Testgruppe 3 könnte bedeuten, dass von den Probanden der Zusammenhang zwischen der Arbeit mit vorgefertigtem Versuchsmaterial und der Herstellung eigener Produkte nicht mehr hergestellt wird. Eine Fehlinterpretation wäre es, wenn die hohe Niveaustufe in der

Experimentierphase 1 mit der Übernahme eines Schematismus erklärt werden würde, weil die vorbereitende Instruktion neben dem vorbereitenden Charakter auch einen informierenden Charakter hat, der besagt, dass die Durchführung von Messungen zum Zweck der Datenaufnahme nicht zu den Inhalten des Vermittlungskonzeptes gehört (vgl. Kap. 3). Bei dem großen Anteil an Probanden, die in den Testgruppen die Niveaustufenkombinationen 3/2 bzw. 3/1 erreicht haben, darf nicht unterschlagen werden, dass an einige Probanden auch die Niveaustufenkombination 1/1 vergeben werden musste (vgl. auch Experimentiertyp D). Besonders auffällig ist hierbei der unsystematische Umgang mit Variablen in der Experimentierphase 1, obwohl die Probanden eine vorbereitende Instruktion erhalten haben. Weiterhin kann für diese Probanden aufgezeigt werden, dass ihre Ergebnisse in einem begleitenden, paper-pencil-basierten Naturwissenschaftlichen Arbeitsweisen Test (Akronym NAW-Test) im unteren Punktebereich liegen. Beide Ergebnisse zusammen (NAW-Test & Beobachtung des Experimentierens) lassen vermuten, dass die Probanden die Variablenkontrollstrategie als Schlüsselkompetenz des systematischen Experimentierens noch nicht hinreichend beherrschen. Die Niveaustufenkombination 1/1 wurde unter den 24 Probanden der Testgruppe 1 zweimal, unter den 10 Probanden der Testgruppe 2 keinmal und unter den 24 Probanden der Testgruppe 3 zweimal vergeben. Ergänzend sei hinzugefügt, dass die Niveaustufenkombination unter den 24 Probanden der Kontrollgruppe sechsmal vergeben wurde, wovon vier Probanden auch im NAW-Test dem unteren Punktebereich zugeordnet werden mussten.

Wie durch den Vergleich des NAW-Tests mit den Niveaustufen des Experimentierens deutlich wird, werden in dieser Arbeit die Ergebnisse verschiedener Testformate (NAW-Test und Beobachtungsbogen) probandenweise trianguliert. Die sich daran anschließenden Korrelationsanalysen ergaben für alle Messzeitpunkte (Vortest, Nachtest, Verzögerter Nachtest) einen positiven signifikanten Zusammenhang mit den Niveaustufen des Experimentierens ($N = 82$ Probanden). Als Korrelationsmaß wurde der Rangkorrelationskoeffizient r_s nach Spearman verwendet, der zwischen der Arbeit mit vorgefertigtem Versuchsmaterial (Experimentierphase 1) und den Ergebnissen des NAW-Tests zu erkennbar höheren Merkmalszusammenhängen führte als zwischen der Herstellung eigener Produkte (Experimentierphase 2) und den Ergebnissen des NAW-Tests. Das Maximum des Korrelationskoeffizienten $r_{s \max}$ konnte für die Arbeit mit vorgefertigtem Versuchsmaterial (Experimentierphase 1) und dem Nachtest bestimmt werden. Mit $r_{s \max} = 0,48$ handelt es sich erwartungsgemäß nicht um einen Merkmalszusammenhang großer Effektstärke (vgl. Kap. 2.2.1 & Kap. 4.5), der einen Austausch der Diagnoseinstrumente und die Aufnahme prozessbezogener Fähigkeiten beim Experimentieren durch schriftliche Testformate rechtfertigen würde.

Den Anlass zum Einsatz des NAW-Tests gaben die beiden Nebenforschungsfragen, mit denen die experimentalgruppenspezifischen Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen festgestellt werden sollen (Forschungsfrage 3). Ein weitergehendes Interesse der Untersuchung liegt darin, zu erfassen, inwieweit die durch die vorbereitenden Instruktionen erworbenen allgemeinen Fähigkeiten im Umgang mit Variablen nachhaltig sind (Forschungsfrage 4). Um beiden Forschungsfragen nachgehen zu können, wurde dem bereits vorhandenen Design aus vorbereitender Instruktion und offener Experimentieraufgabe ein Pre-Post-Follow-Up Design hinzugefügt. Zu allen drei Messzeitpunkten wurde dasselbe Messinstrument eingesetzt. Die erste Datenerhebung mit dem NAW-Test fand unmittelbar vor der vorbereitenden Instruktion (Vortest, Messzeitpunkt 1), die zweite Datenerhebung unmittelbar nach der Bearbeitung der offenen Experimentieraufgabe (Nachtest, Messzeitpunkt 2) und die dritte Datenerhebung im zeitlichen Abstand von drei Monaten zum Nachtest statt (Verzögerter Nachtest, Messzeitpunkt 3). Die Testergebnisse werden mit t-Tests für abhängige Stichproben (Experimentalgruppen) unter Angabe des Effektstärkenmaßes d' – eine Modifikation des Effektstärkenmaßes d nach Cohen für unabhängige Stichproben – ausgewertet. Gegenüber einer Varianzanalyse mit Messwiederholung (RM-ANOVA) bieten die t-Tests den Vorteil, dass die Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen von Messzeitpunkt zu Messzeitpunkt akzentuiert herausgestellt werden können. Zur Wahl der Auswertungsmethoden und den Vorzügen bzw. Fehlerquellen der statistischen Verfahren sei auf das Kapitel 4.3 verwiesen.

Als ein zentrales Ergebnis der Datenauswertung zeigt sich, dass die statistischen Analysen nicht immer zu signifikanten Unterschieden führen, wenn die Messzeitpunkte 1 und 2 sowie 2 und 3 paarweise auf einem Signifikanzniveau von 5 % verglichen werden. Während die Änderungen vom Vor- zum Nachtest und vom Nach- zum Verzögerten Nachtest für die Kontrollgruppe nicht signifikant sind, ist bei den Testgruppen 1 und 3 zwischen den Messzeitpunkten 1 und 2 ein signifikanter Anstieg mit mittlerer Effektstärke festzustellen. Weiterhin stimmen beide Testgruppen darin überein, dass zwischen den Messzeitpunkten 2 und 3 keine signifikanten Unterschiede zu verzeichnen sind. Eine Anomalie der statistischen Datenauswertung liegt in der Testgruppe 2 vor. Die Testgruppe hebt sich von den beiden anderen Experimentalgruppen mit vorbereitender Instruktion ab, weil zwischen den Messzeitpunkten 1 und 2 keine signifikanten Änderungen und zwischen den Messzeitpunkten 2 und 3 signifikante Zuwächse mit hoher Effektstärke festgestellt werden können. Die möglichen Ursachen der Anomalien wurden bereits im Kapitel 4.3 ausführlich diskutiert und sollen an dieser Stelle nicht mehr rezipiert werden. Infolgedessen können die Ergebnisse der Testgruppe 2 bei der Interpretation der statistischen Effekte nicht berücksichtigt werden.

Die Ergebnisse der t-Tests erlauben die Interpretation, dass die signifikanten Änderungen der allgemeinen Fähigkeiten im Umgang mit Variablen auf die Wirkung der einmaligen, einführenden, informierenden, vorbereitenden und expliziten Instruktion zurückgeführt werden kann. Diese Erklärung wird aus einem Kontroll-Testgruppen Vergleich gefolgert, wonach die Bearbeitung der offenen Experimentieraufgabe und die Kombination aus vorbereitender Instruktion und offener Experimentieraufgabe als Ursachen der signifikanten Zuwächse weitestgehend ausgeschlossen werden können. Denn die fast übereinstimmenden statistischen Kennwerte der Kontrollgruppe im Vor- und Nachtest zeigen an, dass die Bearbeitung der offenen Experimentieraufgabe unter dem fehlenden Einfluss der vorbereitenden Instruktion zu keiner signifikanten Änderung führt. Aufgrund der gelungenen zufallsverknüpften Parallelisierung der Experimentalgruppen (KG, TG 1 – 3), die u. a. durch den Einsatz einer einfaktoriellen Varianzanalyse ohne Messwiederholung bestätigt wird, darf weiterhin gefolgert werden, dass auch bei den Probanden der Testgruppen (TG 1, TG 3) die Bearbeitung der offenen Experimentieraufgabe nicht als Grund für die signifikanten Anstiege zwischen den Messzeitpunkten 1 und 2 vereinnahmt werden kann.

Über die kurzfristige Wirkung hinaus, werden die nicht signifikanten Änderungen in den Testgruppen zwischen den Messzeitpunkten 2 und 3 als ein Indikator dafür gewertet, dass die vorbereitenden Instruktionen nachhaltig wirken. In Kombination mit den nicht signifikanten Ergebnissen der Kontrollgruppe darf jedoch auch vermutet werden, dass die Probanden ihre allgemeinen Fähigkeiten im Umgang mit Variablen unter dem Einfluss des schulischen Physikunterrichts nicht verbessern konnten. Der Zeitraum zwischen dem Nachtest und dem Verzögerten Nachtest betrug drei Monate. Eine Beschreibung des erhobenen Konstrukts ist im Kapitel 4.2.2 zu finden.

In einem abschließenden Resümee kann für diese hypothesengenerierende Arbeit konstatiert werden, dass sie einen Beitrag zur Didaktik des Experimentierens leistet. Die beschriebenen prozessbezogenen Fähigkeiten des Experimentierens fußen auf Ergebnissen, die mit der Methode der direkten Beobachtung unter Einsatz eines selbst entworfenen Messinstruments (Beobachtungsbogen) generiert wurden. Zur Aufbereitung der Daten wurden Prozessgrafiken erstellt, die die Identifikation von vier Experimentiertypen (A – D) ermöglichten. Sie schufen die Voraussetzung dafür, die beobachteten Experimentierprozesse in drei empirisch bestimmte Niveaustufen zu überführen und mit den paper-pencil-basierten Ergebnissen des NAW-Tests von drei Messzeitpunkten vergleichen zu können. Weiterhin versteht sich diese Arbeit als vorbereitende Studie zur Erstellung eines nicht normativen Kompetenzentwicklungsmodells, das auf Basis von beobachtbaren Experimentierprozessen beschrieben wird. Als potentielle Entwicklungsstufen der naturwissenschaftlichen Arbeitsweise werden die Unter-, Mittel- und Oberstufe propagiert. Im

Rahmen des Vorschlags werden mit dieser Studie die Fähigkeiten im Umgang mit Variablen beim Experimentieren erfasst, die Gymnasiasten am Ende der SI zeigen. Ein bis dato unbeantwortetes Forschungsinteresse liegt darin, zu erfassen, wie nachhaltig die durch die vorbereitende Instruktion erworbenen Fähigkeiten im Umgang mit Variablen bei der Bearbeitung einer offenen Experimentieraufgabe sind. Zudem besteht ein Entwicklungsbedarf an geeigneten Unterstützungsmaßnahmen, mit denen der unter wechselnden Experimentierbedingungen festgestellte Niveauabfall kompensiert werden kann. Obwohl die Forschungsanliegen momentan noch nicht beantwortet werden können, geben die Ergebnisse für die Vermittlung des Variablenkonzeptes Anlass zu der Aussage: Mehr Offenheit beim Experimentieren wagen!

Anhang

Anhangsverzeichnis

A.1 Materialien und Medien der vorbereitenden Instruktionen	173
A.1.1 Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Reproduktion (TG, TG 1)	173
A.1.2 Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Reorganisation (TG 2)	174
A.1.3 Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Transfer (TG 3).....	176
A.1.4 Experimentieranordnungen zur Aufnahme von Messfolgen für die vorbereitenden Instruktionen und der offenen Experimentieraufgabe.....	178
A.1.5 Präsentation für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Reproduktion (TG, TG 1).....	179
A.1.6 Arbeitsblatt für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Reproduktion (TG, TG 1).....	185
A.1.7 Präsentation für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Reorganisation (TG 2).....	187
A.1.8 Arbeitsblatt für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Reorganisation (TG 2).....	193
A.1.9 Präsentation für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Transfer (TG 3)	195
A.1.10 Arbeitsblatt für die vorbereitende Instruktion der Testgruppen mit der Transferleistung Transfer (TG 3)	201
A.2 Messinstrumente	203
A.2.1 Fach- & NAW-Test	203
A.2.2 Beobachtungsbogen.....	210
A.3 Ergebnisse	225
A.3.1 Verteilung der Ergebnisse des NAW-Tests zum Messzeitpunkt 1 der Probandenpaare der Experimentalgruppen	225
A.3.2 Vergleich der Ergebnisse des NAW-Tests zu den Messzeitpunkten 1 – 3 mit den Niveaustufen des Experimentierens.....	226

A.1.1

Bestimmung maximaler Drehzahlen und Spannungen U für Windradprofile mit rechteckiger Grundform in Abhängigkeit von Testvariablen			
Testvariable	Ausprägung	U_{\max} in min^{-1}	U_{\max} in V
Flügelbreite	1,0 cm	1000	0,81
	1,5 cm	1220	0,99
	2,0 cm	1410	1,15
	2,5 cm	1640	1,33
	3,0 cm	1860	1,51
Anstellwinkel	5 °	2250	1,82
	10 °	2460	1,96
	15 °	2330	1,89
	20 °	1850	1,51
	25 °	1550	1,26
	30 °	1280	1,01
	35 °	1100	0,88
	45 °	940	0,76
	60 °	630	0,52
Flügelanzahl	1	700	0,57
	2	1220	0,99
	3	1410	1,14
	4	1420	1,15
	5	1400	1,14
	6	1350	1,09
Kontrollvariablen & Ausgangsparameter	Anmerkung: Flügelbreite (2,0 cm), Anstellwinkel (22 °), Flügelanzahl (2), Flügellänge (7,0 cm), Profillänge (14,0 cm), Winglets (0), Wölbung (0), Verdrehung bzw. Verschränkung (0), Material (PVC)		
Profilanströmung	Haartrockner		
Geschwindigkeit	Betrag der Profilanströmungsgeschwindigkeit: $v = 25 \text{ km/h}$ bzw. 7 m/s		
Messgröße	Spannung und Drehzahl (berührungsloser Drehzahlmesser RC 200 von Jamara)		

Tab. A 1: Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppen TG (Vorstudie) und TG 1 (Hauptstudie) mit der Transferleistung
Reproduktion

A.1.2

Bestimmung maximaler Auftriebskräfte F_{\max} für Tragflächenprofile mit planar-rechteckiger Grundform in Abhängigkeit von Testvariablen		
Testvariable	Ausprägung	F_{\max} in cN
Spannweite	4,0 cm	5,2
	6,0 cm	6,8
	8,0 cm	9,2
	10,0 cm	11,5
	12,0 cm	13,9
	14,0 cm	15,8
Anzahl Querrillen auf Profilunterseite	0	11,3
	1	11,1
	2	11,5
	3	11,7
	4	11,3
	5	11,5
	6	11,3
	7	11,5
	8	11,4
	9	11,3
	10	11,0
Kontrollvariablen & Ausgangsparameter	Anmerkung: Spannweite (10,0 cm), Anstellwinkel (15 °), Flügelanzahl (1), Flügelbreite (7,0 cm), Winglets (0), Masse (5,5 g) Anzahl Querrillen auf Flügelunterseite (0), Material (Depron)	
Profilanströmung	Ventilator Typ KDC 306 von Siemens-Schuckertwerke AG	
Geschwindigkeit	Betrag der Profilanströmgeschwindigkeit: $v = 5,5 \text{ m/s}$	
Messvorrichtung	Digitale Waage von Sartorius	
Messgröße	maximale Massenabnahme bzw. Massenzunahme Δm_{\max}	

Tab. A 2a: Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 2 (Hauptstudie) mit der Transferleistung Reorganisation

Fortsetzung A.1.2

Fortsetzung Bestimmung maximaler Auftriebskräfte F_{\max} für Tragflächenprofile mit planar-rechteckiger Grundform in Abhängigkeit von Testvariablen		
Testvariable	Ausprägung	F_{\max} in cN
Anstellwinkel	- 35 °	- 8,6
	- 30 °	- 7,6
	- 25 °	- 6,8
	- 20 °	- 4,7
	- 15 °	- 2,6
	- 10 °	- 1,6
	- 5 °	+ 2,1
	0 °	+ 3,7
	+ 5 °	+ 6,3
	+ 10 °	+ 8,3
	+ 15 °	+ 10,5
	+ 20 °	+ 12,4
	+ 25 °	+ 13,1
	+ 30 °	+ 13,7
	+ 35 °	+ 14,2
	+ 40 °	+ 14,3
	+ 45 °	+ 14,2
	+ 50 °	+ 13,6
	+ 55 °	+ 11,4
	+ 60 °	+ 10,2
	+ 65 °	+ 8,6
	+ 70 °	+ 6,7
	+ 75 °	+ 3,9
Kontrollvariablen & Ausgangsparameter	Anmerkung: Spannweite (10,0 cm), Anstellwinkel (15 °), Flügelanzahl, (1), Flügelbreite (7,0 cm), Winglets (0), Masse (5,5 g), Anzahl Querrillen auf Flügelunterseite (0), Material (Depron)	
Profilanströmung	Ventilator Typ KDC 306 von Siemens-Schuckertwerke AG	
Geschwindigkeit	Betrag der Profilanströmgeschwindigkeit: $v = 5,5 \text{ m/s}$	
Messvorrichtung	Digitale Waage von Sartorius	
Messgröße	maximale Massenabnahme bzw. Massenzunahme Δm_{\max}	

Tab. A 2b: Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 2 (Hauptstudie) mit der Transferleistung Reorganisation

A.1.3

Bestimmung maximaler Wurfweiten s_{\max} für Wurfkörper mit kugeliger Grundform in Abhängigkeit von Testvariablen		
Testvariable	Ausprägung	s_{\max} in cm
Masse	13,0 g	755
	14,0 g	788
	15,0 g	840
	20,0 g	831
	25,0 g	822
	30,0 g	790
	35,0 g	756
	40,0 g	723
	45,0 g	690
	50,0 g	655
	55,0 g	624
Durchmesser	3,0 cm	1200
	4,0 cm	1160
	5,0 cm	1163
	6,0 cm	1157
	7,0 cm	1125
	8,0 cm	958
	10,0 cm	795
	12,0 cm	704
	14,0 cm	512
Kontrollvariablen & Ausgangsparameter	Anmerkung: Kugelmasse (14,0 g), Kugeldurchmesser (10,0 cm), Oberfläche bzw. Anzahl Löcher in Kugeloberfläche (0), Kugelmaterial (Styropor)	
Messanordnung		
	Katapult (Eigenbau, Abb. A 3) mit der Federhärte $D = 2,2 \text{ N/cm}$	

Tab. A 3a: Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 3 (Hauptstudie) mit der Transferleistung Transfer

Fortsetzung A.1.3

Fortsetzung Bestimmung maximaler Wurfweiten s_{\max} für Wurfkörper mit kugeliger Grundform in Abhängigkeit von Testvariablen		
Testvariable	Ausprägung	s_{\max} in cm
Anzahl Löcher in Kugeloberfläche	0	780
	10	805
	20	814
	30	804
	40	786
	50	781
	60	769
	70	743
	80	744
	90	744
	100	735
	120	730
	140	720
	160	706
Kontrollvariablen & Ausgangsparameter	Anmerkung: Kugelmasse (14,0 g), Kugeldurchmesser (10,0 cm), Oberfläche bzw. Anzahl Löcher in Kugeloberfläche (0), Kugelmaterial (Styropor)	
Messanordnung	Katapult (Eigenbau, Abb. A 3) mit der Federhärte $D = 2,2 \text{ N/cm}$	

Tab. A 3b: Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 3 (Hauptstudie) mit der Transferleistung Transfer

A.1.4

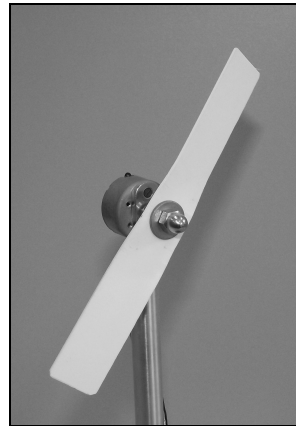


Abb. A 1: Experimentieranordnung zur Anströmung von Windradprofilen zur Aufnahme von Messfolgen für die vorbereitende Instruktion der Testgruppen TG (Vorstudie) und TG 1 (Hauptstudie) mit der Transferleistung Reproduktion sowie zur Bearbeitung der offenen Experimentieraufgabe

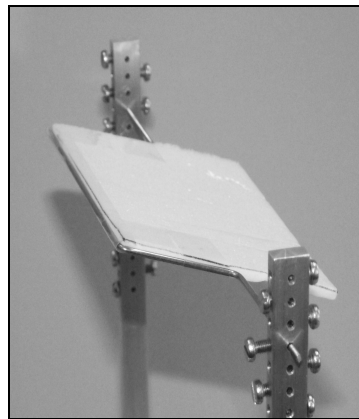
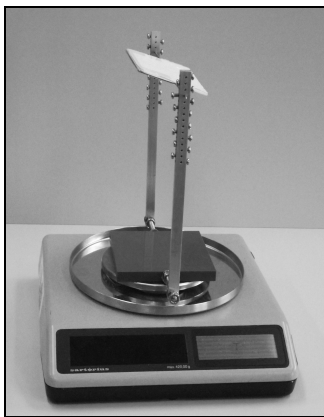


Abb. A 2: Experimentieranordnung zur Anströmung von Tragflächenprofilen zur Aufnahme von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 2 (Hauptstudie) mit der Transferleistung Reorganisation

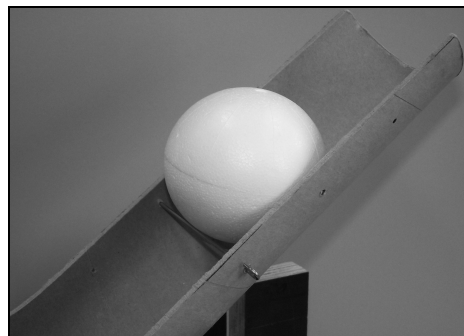
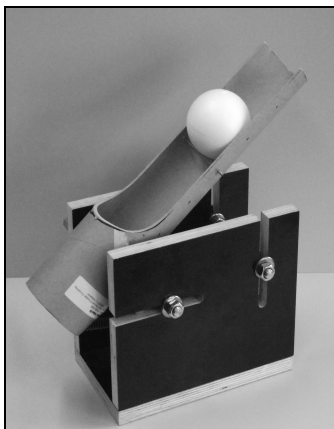


Abb. A 3: Experimentieranordnung zum Abwurf von Wurfkörpern zur Aufnahme von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 3 (Hauptstudie) mit der Transferleistung Transfer

A.1.5

RUHR-UNIVERSITÄT BOCHUM

Universität

FAKULTÄT FÜR PHYSIK UND ASTRONOMIE
DIDAKTIK DER PHYSIK

Stefan Kirchner

Der Umgang mit Variablen beim Experimentieren
am Beispiel der Profilkonstruktion
für ein Windradmodell

Informierende Vorbereitung zum Experimentieren

KIRCHNER FOLIE 1

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Fragestellung

Wie müssen
die Flügel eines Windrades
konstruiert bzw. geformt werden,
damit es sich möglichst schnell dreht?

KIRCHNER FOLIE 2

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Beispiele möglicher Flügelformen von Windrädern, ...



... die unterschiedliche Aufgaben erfüllen !

KIRCHNER FOLIE 3

Fortsetzung A.1.5

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>Konkretisierung der Fragestellung ...</h3> <p>Wie muss man beim Experimentieren vorgehen, damit man ein möglichst schnell drehendes Windrad erhält?</p>	
KIRCHNER	FOLIE 4

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>... führt zum Thema</h3> <p>Der Umgang mit Variablen beim Experimentieren</p> <p style="text-align: center;">↓</p> <ul style="list-style-type: none"> • Welche Variablen haben einen Einfluss auf die Drehzahl ? (Variablen identifizieren & suchen, Planung einer Testreihe) • Wie ändert sich die Drehzahl bei system. Änderung einer Variablen ? (Verlauf & Effektstärken der Variablen innerhalb einer Testreihe, Bewertung) • Wie stark ist der Einfluss einer Variablen auf die Drehzahl ? (Gewichtung von Variablen, Vergleich & Bewertung von Testreihen) 	
KIRCHNER	FOLIE 5

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>Gliederung</h3> <ul style="list-style-type: none"> • <u>Einführung:</u> Begriffserklärungen, Variablen identifizieren, Ausgangsbeispiel • <u>Planung von Experimenten:</u> Experimentierplan mit Test- & Kontrollvariablen • <u>Umsetzung des Experimentierplans:</u> Experimentieranordnung & -demonstration zur Aufnahme der Messdaten • <u>Auswertung der realen Messdaten:</u> Vergleich & Bewertung von Variablen mittels Arbeitsblatt • <u>Abschluss:</u> Variablen suchen, Vertiefung, Zusammenfassung & Reflexion 	
KIRCHNER	FOLIE 6

Fortsetzung A.1.5

RUHR-UNIVERSITÄT BOCHUM DIDAKTIK DER PHYSIK

Einführung

Variablen:
Einflussgrößen, die beim Experimentieren verändert werden

Unterscheidung der Variablen:

- Variablen der Experimentierbedingungen, z.B.
Windgeschwindigkeit, Luftdichte, Temperatur, Standort, ...
- Variablen der Windradflügel, z.B.
Flügelbreite, Anstellwinkel, Flügelanzahl, Flügellänge, ...

➔ hier: Beschränkung auf Variablen der Windradflügel !!!

KIRCHNER FOLIE 7

RUHR-UNIVERSITÄT BOCHUM DIDAKTIK DER PHYSIK

Einführung

Ausgangsbeispiel

Wahl eines Windrades
mit rechteckiger Grundform
und den Ausgangsparametern

Breite; Anstellwinkel; Anzahl; Länge
(2,0 cm; 22° ; 2 ; 7,0 cm)

KIRCHNER FOLIE 8

RUHR-UNIVERSITÄT BOCHUM DIDAKTIK DER PHYSIK

Einführung

Abb.
Ausgangsbeispiel


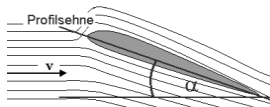


Abb.
Anstellwinkel



Mögliche Fragestellungen zur Untersuchung von Variablen

- Hat die Flügelbreite einen Einfluss auf die Drehzahl ?
- Hat der Anstellwinkel einen Einfluss auf die Drehzahl ?
- Hat die Flügelanzahl einen Einfluss auf die Drehzahl ?
- ...

KIRCHNER FOLIE 9

Fortsetzung A.1.5

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Flügelbreite

	<u>Breite</u>	<u>Winkel</u>	<u>Anzahl</u>	<u>Länge</u>
Flügel	(: :)	:	:	:
Flügel	(1,0 cm	22°	2	7,0 cm)
Flügel	(2,0 cm	22°	2	7,0 cm)
Flügel	(3,0 cm	22°	2	7,0 cm)
Flügel	(4,0 cm	22°	2	7,0 cm)
Flügel	(: :)	:	:	:

- Experimentierplan, der vergleichende Untersuchung zum Test der Variablen Flügelbreite ermöglicht

KIRCHNER

FOLIE 10

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Flügelbreite

Testvariable:

Variable, die verändert wird

	<u>Breite</u>	<u>Winkel</u>	<u>Anzahl</u>	<u>Länge</u>
Flügel	(: :)	:	:	:)
Flügel	(1,0 cm	22°	2	7,0 cm)
Flügel	(2,0 cm	22°	2	7,0 cm)
Flügel	(3,0 cm	22°	2	7,0 cm)
Flügel	(4,0 cm	22°	2	7,0 cm)
Flügel	(: :)	:	:	:)

KIRCHNER

FOLIE 11

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Flügelbreite

Kontrollvariablen:

Variablen, die nicht verändert werden

Breite

Winkel

Anzahl

Länge

Flügel	(:	:	:	:)
Flügel	(1,0 cm	22°	2	7,0 cm)
Flügel	(2,0 cm	22°	2	7,0 cm)
Flügel	(3,0 cm	22°	2	7,0 cm)
Flügel	(4,0 cm	22°	2	7,0 cm)
Flügel	(:	:	:	:)

KIRCHNER

FOLIE 12

Fortsetzung A.1.5

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Flügelbreite

	<u>Breite</u>	<u>Winkel</u>	<u>Anzahl</u>	<u>Länge</u>
Flügel	(:)	(:)	(:)	(:)
Flügel	(1,0 cm)	(22°)	(2)	(7,0 cm)
Flügel	(2,0 cm)	(22°)	(2)	(7,0 cm)
Flügel	(3,0 cm)	(22°)	(2)	(7,0 cm)
Flügel	(4,0 cm)	(22°)	(2)	(7,0 cm)
Flügel	(:)	(:)	(:)	(:)

Dimensionen der Variablen

KIRCHNER

FOLIE 13

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Anstellwinkel

Testvariable

Breite

Winkel

Anzahl

Länge

Flügel	(:	:	:	:)
Flügel	(2,0 cm	5°	2	7,0 cm)
Flügel	(2,0 cm	10°	2	7,0 cm)
Flügel	(2,0 cm	15°	2	7,0 cm)
Flügel	(2,0 cm	20°	2	7,0 cm)
Flügel	(:	:	:	:)

Kontrollvariablen

KIRCHNER

FOLIE 14

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan

	<u>Breite</u>	<u>Winkel</u>	<u>Anzahl</u>	<u>Länge</u>
Flügel	(:)	(:)	(:)	(:)
Flügel	(2,0 cm)	(5°)	(2)	(7,0 cm)
Flügel	(3,0 cm)	(10°)	(2)	(7,0 cm)
Flügel	(2,0 cm)	(15°)	(3)	(7,0 cm)
Flügel	(2,0 cm)	(20°)	(2)	(8,0 cm)
Flügel	(:)	(:)	(:)	(:)



**Warum ist eine vergleichende Untersuchung
zum Test einer Variablen mit diesem
Experimentierplan nicht möglich?**

KIRCHNER

FOLIE 15

Fortsetzung A.1.5

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Abschluss: Zusammenfassung und Reflexion			
Variablen identifizieren, Variablen suchen	Var. 1; Var. 2; Var. 3; ... Bsp.: Breite; Winkel; Anzahl; ...		
Variablen testen, und kontrollieren	Testvar. 2 (Bsp. Winkel) Kontrollvar. 1; 3; ... (Bsp.: Breite; Anzahl; ...)		
Variablen vergleichen, Variablen bewerten	Gewichtung (Var 2 > Var 1 > Var 3 ...) Bsp. (Winkel > Breite > Anzahl ...) Effektstärke (Var. 1, 2 > Var. 3 ...) Bsp. (Breite, Winkel > Anzahl ...)		
KIRCHNER		FOLIE 16	

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Abschluss: Zusammenfassung und Reflexion			
Variablen identifizieren, Variablen suchen	 	ist eine Voraussetzung für ist eine Voraussetzung für	
Variablen testen, und kontrollieren			
Variablen vergleichen, Variablen bewerten			
KIRCHNER		FOLIE 17	

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Vielen Dank für Eure Aufmerksamkeit			
KIRCHNER		FOLIE 18	

A.1.6

Messung maximaler Drehzahlen von Windradprofilen**1. Ausgangsparameter eines Windradprofils mit rechteckiger Grundform:**

Flügelbreite (2,0 cm), Anstellwinkel (22°), Flügelanzahl (2), Flügellänge (7,0 cm)

2. Messfolgen zum Test der Variablen

a) *Flügelbreite*, b) *Anstellwinkel* und c) *Flügelanzahl* mit der Anströmgeschwindigkeit $v = 25 \text{ km/h}$

2 a)		2 b)		2 c)	
<i>Flügelbreite</i> in cm	max. Drehzahl in U/ min	<i>Anstellwinkel</i> in °	max. Drehzahl in U/ min	<i>Flügelanzahl</i>	max. Drehzahl in U/ min
⋮	⋮	⋮	⋮	⋮	⋮
1,0	1000	5	2250	2	1220
1,5	1220	10	2460	3	1410
2,0	1410	15	2330	4	1420
2,5	1640	20	1850	5	1400
3,0	1860	25	1550	6	1350
⋮	⋮	⋮	⋮	⋮	⋮
<u>Testvariable:</u> <i>Flügelbreite</i>		<u>Kontrollvariablen:</u> Anstellwinkel (22°) Flügelanzahl (2) Flügellänge (7,0 cm)		<u>Testvariable:</u> <i>Flügelanzahl</i>	
		<u>Kontrollvariablen:</u> Flügelbreite (2,0 cm) Flügelanzahl (2) Flügellänge (7,0 cm)		<u>Kontrollvariablen:</u> Flügelbreite (2,0 cm) Anstellwinkel (22°) Flügellänge (7,0 cm)	

Fortsetzung A.1.6

3. Graphische Auswertung der Messfolgen

(Anströmgeschwindigkeit $v = 25 \text{ km/h}$)

Diagramm 1

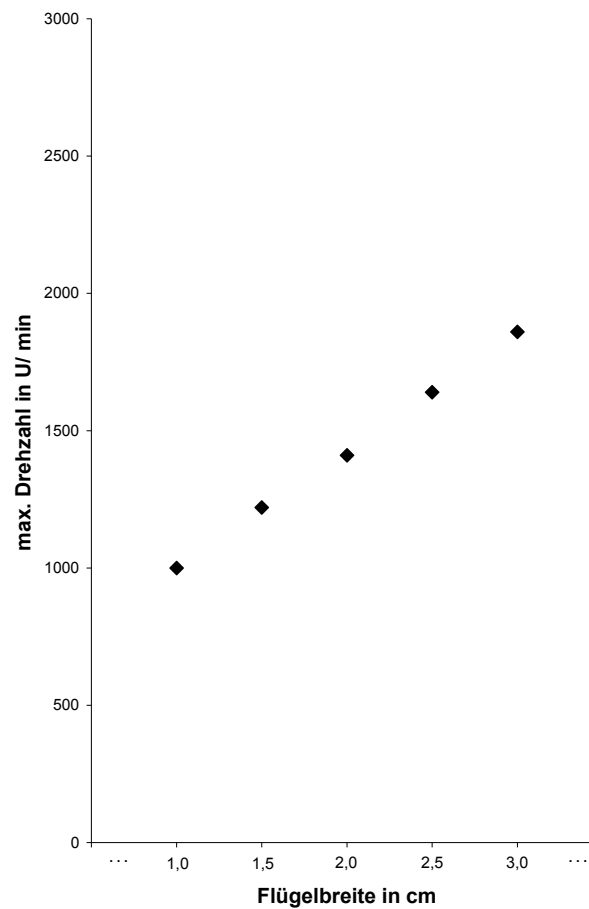


Diagramm 2

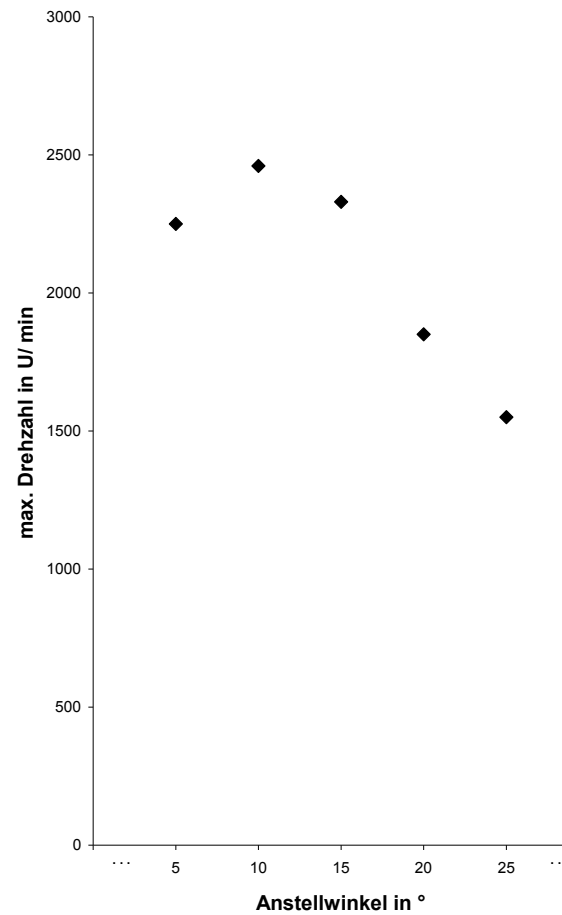


Diagramm 3

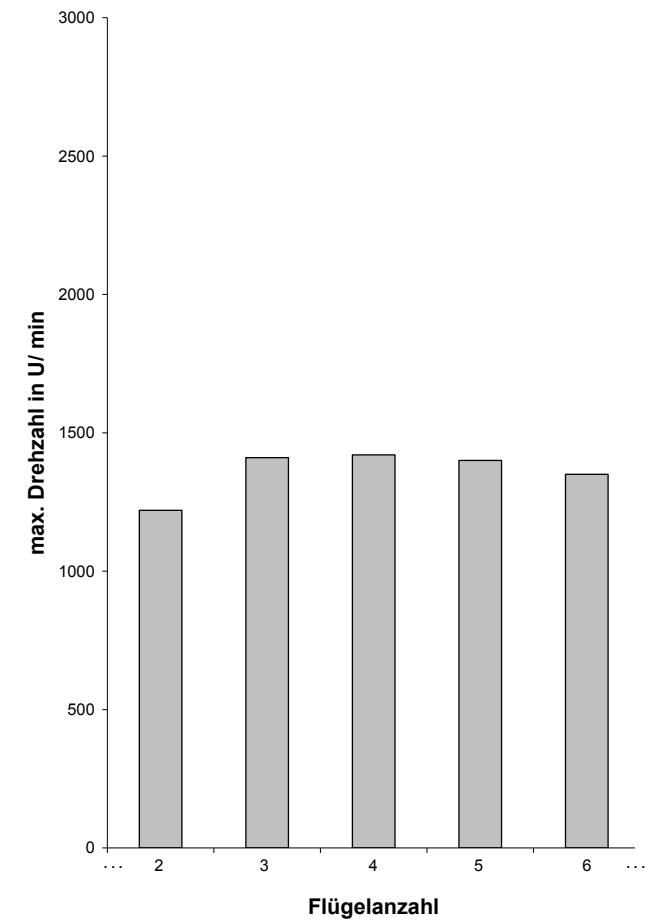


Diagramme: Abhängigkeit der maximalen Drehzahl von der Flügelbreite (1), vom Anstellwinkel (2) und von der Flügelanzahl (3) des Windradprofils

A.1.7

RUHR-UNIVERSITÄT BOCHUM

Universität

FAKULTÄT FÜR PHYSIK UND ASTRONOMIE
DIDAKTIK DER PHYSIK

Stefan Kirchner

Der Umgang mit Variablen beim Experimentieren
am Beispiel der Tragflächenkonstruktion
für ein Flugzeugmodell

Informierende Vorbereitung zum Experimentieren

KIRCHNER FOLIE 1

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Fragestellung

Wie müssen
die Tragflächen eines Flugzeuges
konstruiert bzw. geformt werden,
damit eine möglichst große Auftriebskraft wirkt?

KIRCHNER FOLIE 2

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Beispiele möglicher Flügelformen, ...



... die unterschiedliche Aufgaben erfüllen !

KIRCHNER FOLIE 3

Fortsetzung A.1.7

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<p>Konkretisierung der Fragestellung ...</p> <p>Wie muss man beim Experimentieren vorgehen, damit man ein Tragflächenprofil erhält, dass eine möglichst große Auftriebskraft erfährt?</p>	
KIRCHNER	FOLIE 4


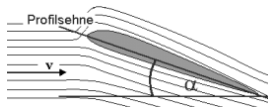
RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<p>... führt zum Thema</p> <p>Der Umgang mit Variablen beim Experimentieren</p> <p>↓</p> <ul style="list-style-type: none"> • Welche Variablen haben einen Einfluss auf die Auftriebskraft ? (Variablen identifizieren & suchen, Planung einer Testreihe) • Wie ändert sich die Auftriebskraft bei system. Änderung einer Variablen ? (Verlauf & Effektstärken der Variablen innerhalb einer Testreihe, Bewertung) • Wie stark ist der Einfluss einer Variablen auf die Auftriebskraft ? (Gewichtung von Variablen, Vergleich & Bewertung von Testreihen) 	
KIRCHNER	FOLIE 5

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<p>Gliederung</p> <ul style="list-style-type: none"> • <u>Einführung:</u> Begriffserklärungen, Variablen identifizieren, Ausgangsbeispiel • <u>Planung von Experimenten:</u> Experimentierplan mit Test- & Kontrollvariablen • <u>Umsetzung des Experimentierplans:</u> Experimentieranordnung & -demonstration zur Aufnahme der Messdaten • <u>Auswertung der realen Messdaten:</u> Vergleich & Bewertung von Variablen mittels Arbeitsblatt • <u>Abschluss:</u> Variablen suchen, Vertiefung, Zusammenfassung & Reflexion 	
KIRCHNER	FOLIE 6

Fortsetzung A.1.7

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
Einführung	
<p><u>Variablen:</u> Einflussgrößen, die beim Experimentieren verändert werden</p> <p><u>Unterscheidung der Variablen:</u></p> <ul style="list-style-type: none"> • Variablen der Experimentierbedingungen, z.B. <i>Windgeschwindigkeit, Luftdichte, Temperatur, ...</i> • Variablen der Tragflächen, z.B. <i>Spannweite, Anstellwinkel, Oberfläche, Flügelbreite, ...</i> <p>➔ hier: Beschränkung auf Variablen der Tragflächen !!!</p>	
KIRCHNER	FOLIE 7

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
Einführung	
<u>Ausgangsbeispiel</u>	
<p>Wahl eines flachen Tragflächenprofils mit rechteckiger Grundform und den Ausgangsparametern</p> <p><i>Spannweite; Anstellwinkel; Oberfläche; Flügelbreite</i> (10,0 cm ; 15° ; glatt ; 7,0 cm)</p>	
KIRCHNER	FOLIE 8

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
Einführung	
<p>Abb. Ausgangsbeispiel</p> 	<p>Abb. Anstellwinkel</p> 
<p><u>Mögliche Fragestellungen zur Untersuchung von Variablen</u></p> <ul style="list-style-type: none"> • Hat die Spannweite einen Einfluss auf die Auftriebskraft ? • Hat der Anstellwinkel einen Einfluss auf die Auftriebskraft ? • Hat die Beschaffenheit der Oberfläche einen Einfluss auf die Auftriebskraft ? • ... 	
KIRCHNER	FOLIE 9

Fortsetzung A.1.7

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Spannweite

	<u>Spannweite</u>	<u>Winkel</u>	<u>Oberfläche</u>	<u>Breite</u>	
Flügel	(:)	:	:	:)
Flügel	(4,0 cm	15°	glatt	7,0 cm)
Flügel	(6,0 cm	15°	glatt	7,0 cm)
Flügel	(8,0 cm	15°	glatt	7,0 cm)
Flügel	(10,0 cm	15°	glatt	7,0 cm)
Flügel	(12,0 cm	15°	glatt	7,0 cm)
Flügel	(:)	:	:	:)

- Experimentierplan, der vergleichende Untersuchung zum Test der Variablen Spannweite ermöglicht

KIRCHNER

FOLIE 10

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Spannweite

Testvariable:

Variable, die verändert wird

	<u>Spannweite</u>	<u>Winkel</u>	<u>Oberfläche</u>	<u>Breite</u>	
Flügel	(:)	:	:	:)
Flügel	(4,0 cm	15°	glatt	7,0 cm)
Flügel	(6,0 cm	15°	glatt	7,0 cm)
Flügel	(8,0 cm	15°	glatt	7,0 cm)
Flügel	(10,0 cm	15°	glatt	7,0 cm)
Flügel	(12,0 cm	15°	glatt	7,0 cm)
Flügel	(:)	:	:	:)

KIRCHNER

FOLIE 11

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Spannweite

Kontrollvariablen:

Variablen, die nicht verändert werden

Spannweite

Winkel

Oberfläche

Breite

Flügel	(:	:	:	:)
Flügel	(4,0 cm	15°	glatt	7,0 cm)
Flügel	(6,0 cm	15°	glatt	7,0 cm)
Flügel	(8,0 cm	15°	glatt	7,0 cm)
Flügel	(10,0 cm	15°	glatt	7,0 cm)
Flügel	(12,0 cm	15°	glatt	7,0 cm)
Flügel	(:	:	:	:)

KIRCHNER

FOLIE 12

Fortsetzung A.1.7

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Spannweite

	<u>Spannweite</u>	<u>Winkel</u>	<u>Oberfläche</u>	<u>Breite</u>
Flügel (:	:	:	:)
Flügel (4,0 cm	15°	glatt	7,0 cm)
Flügel (6,0 cm	15°	glatt	7,0 cm)
Flügel (8,0 cm	15°	glatt	7,0 cm)
Flügel (10,0 cm	15°	glatt	7,0 cm)
Flügel (12,0 cm	15°	glatt	7,0 cm)
Flügel (:	:	:	:)

Dimensionen der Variablen

KIRCHNER

FOLIE 13

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Anstellwinkel

Testvariable

Spannweite

Winkel

Oberfläche

Breite

Flügel	(:	:	:	:)
Flügel	(10,0 cm	0°	glatt	7,0 cm)
Flügel	(10,0 cm	10°	glatt	7,0 cm)
Flügel	(10,0 cm	20°	glatt	7,0 cm)
Flügel	(10,0 cm	30°	glatt	7,0 cm)
Flügel	(10,0 cm	40°	glatt	7,0 cm)
Flügel	(:	:	:	:)

Kontrollvariablen


KIRCHNER

FOLIE 14

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan



	<u>Spannweite</u>	<u>Winkel</u>	<u>Oberfläche</u>	<u>Breite</u>
Flügel	(:)	(:)	(:)	(:)
Flügel	(10,0 cm)	(0°)	(glatt)	(7,0 cm)
Flügel	(8,0 cm)	(10°)	(glatt)	(7,0 cm)
Flügel	(12,0 cm)	(20°)	(glatt)	(7,0 cm)
Flügel	(10,0 cm)	(30°)	(6 Rillen)	(7,0 cm)
Flügel	(10,0 cm)	(40°)	(glatt)	(4,0 cm)
Flügel	(:)	(:)	(:)	(:)



**Warum ist eine vergleichende Untersuchung
zum Test einer Variablen mit diesem
Experimentierplan nicht möglich?**

KIRCHNER

FOLIE 15

Fortsetzung A.1.7

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Abschluss: Zusammenfassung und Reflexion			
Variablen identifizieren, Variablen suchen	Var. 1; Var. 2; Var. 3; ... Bsp.: Spannweite; Anstellwinkel; Oberfläche; ...		
Variablen testen, und kontrollieren	Testvar. 2 (Bsp. Anstellwinkel) Kontrollvar. 1; 3;... (Bsp.: Spannweite; Oberfläche;...)		
Variablen vergleichen, Variablen bewerten	Gewichtung (Var 1 > Var 2 > Var 3 ...) Bsp. (Spannweite > Anstellwinkel > Oberfläche ...) Effektstärke (Var. 1, 2 > Var. 3 ...) Bsp. (Spannweite, Anstellwinkel > Oberfläche ...)		
KIRCHNER		FOLIE 16	

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Abschluss: Zusammenfassung und Reflexion			
Variablen identifizieren, Variablen suchen	 	ist eine Voraussetzung für ist eine Voraussetzung für	
Variablen testen, und kontrollieren			
Variablen vergleichen, Variablen bewerten			
KIRCHNER		FOLIE 17	

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
<p>Vielen Dank für Eure Aufmerksamkeit</p>			
KIRCHNER		FOLIE 18	

A.1.8

Messung maximaler Auftriebskräfte von Tragflächenprofilen**1. Ausgangsparameter eines flachen Tragflächenprofils mit rechteckiger Grundform:**

Spannweite (10,0 cm), Anstellwinkel (15°), Oberfläche (glatt), Flügelbreite (7,0 cm)

2. Messfolgen zum Test der Variablena) *Spannweite*, b) *Anstellwinkel* und c) *Oberfläche* (Anzahl Querrillen auf Flügelunterseite) mit der Anströmgeschwindigkeit $v = 20 \text{ km/h}$

2 a)		2 b)		2 c)	
<i>Spannweite</i> in cm	max. Auftriebskraft in cN	<i>Anstellwinkel</i> in °	max. Auftriebskraft in cN	<i>Anzahl Rillen</i>	max. Auftriebskraft in cN
⋮	⋮	⋮	⋮	0 (glatt)	11,3
4,0	5,2	0	3,7	2	11,5
6,0	6,8	10	8,3		
8,0	9,2	20	12,4	4	11,3
10,0	11,5	30	13,7	6	11,3
12,0	13,9	40	14,3	8	11,4
14,0	15,8	50	13,6	10	11,0
⋮	⋮	⋮	⋮	⋮	⋮
<u>Testvariable:</u> <i>Spannweite</i>		<u>Kontrollvariablen:</u> Anstellwinkel (15°) Oberfläche (glatt) Flügelbreite (7,0 cm)		<u>Testvariable:</u> <i>Oberfläche</i>	
		<u>Kontrollvariablen:</u> Spannweite (10,0 cm) Oberfläche (glatt) Flügelbreite (7,0 cm)		<u>Kontrollvariablen:</u> Spannweite (10,0 cm) Anstellwinkel (15°) Flügelbreite (7,0 cm)	

Fortsetzung A.1.8

3. Graphische Auswertung der Messfolgen

(Anströmgeschwindigkeit $v = 20 \text{ km/h}$)

Diagramm 1

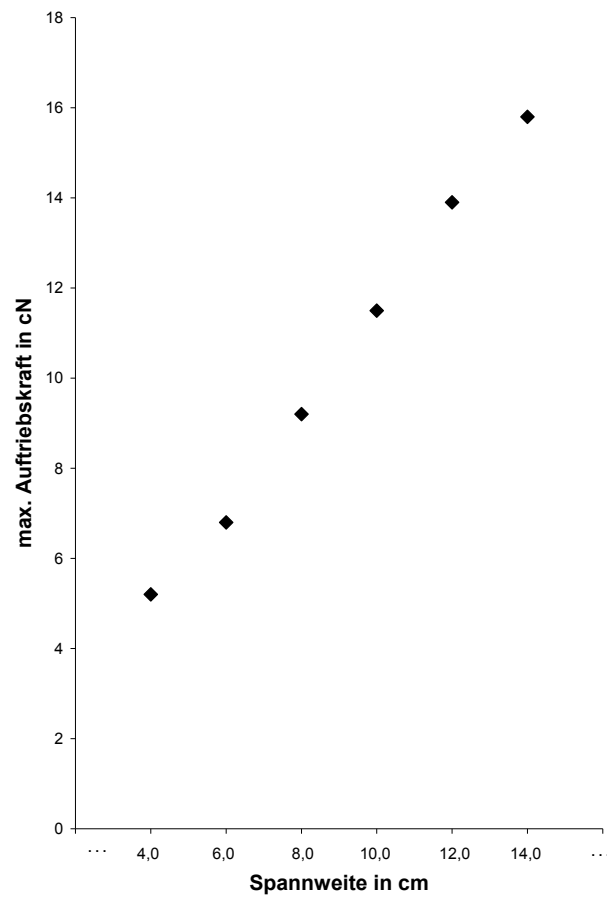


Diagramm 2

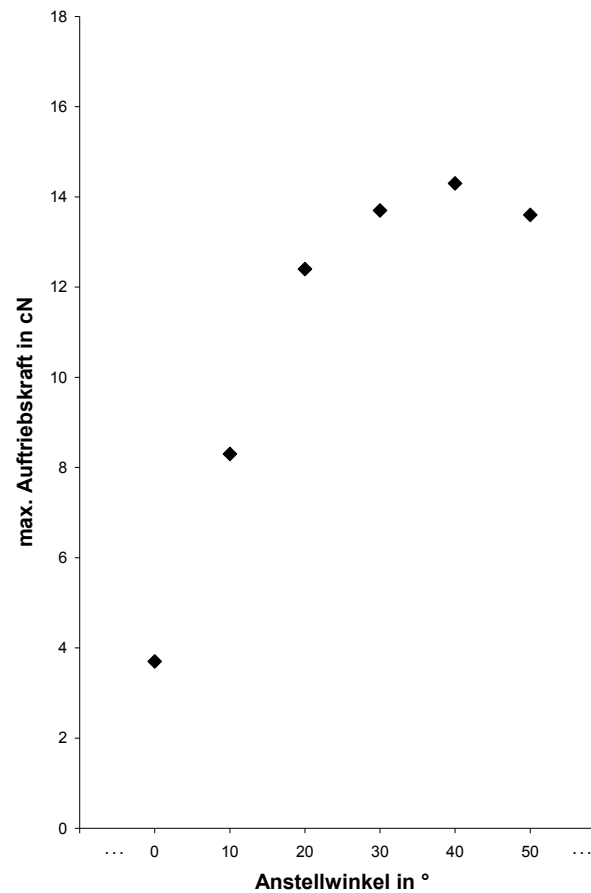


Diagramm 3

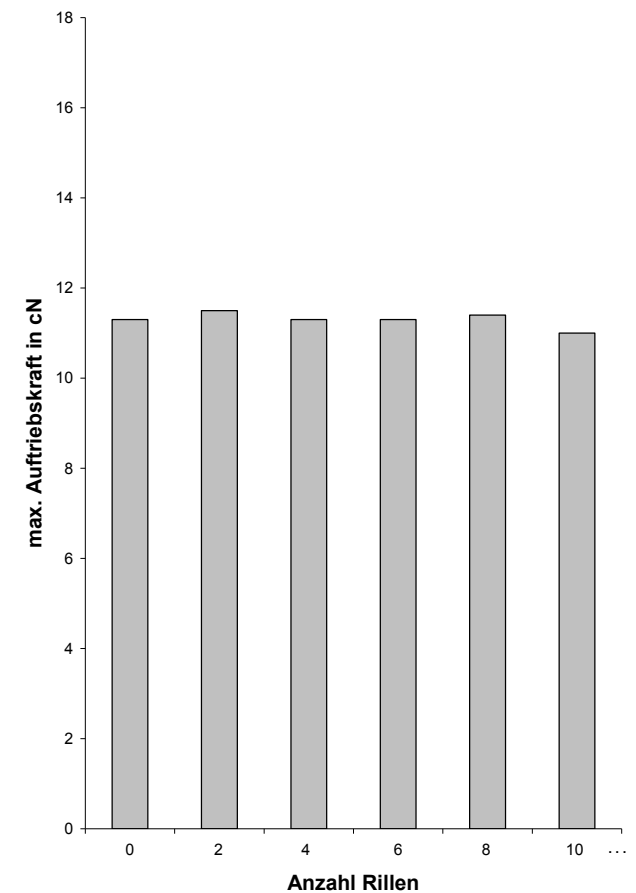


Diagramme: Abhängigkeit der maximalen Auftriebskraft von der Spannweite (1), vom Anstellwinkel (2) und von der Anzahl der Querrillen auf der Flügelunterseite (3) des Tragflächenprofils

A.1.9

RUHR-UNIVERSITÄT BOCHUM

Universität

FAKULTÄT FÜR PHYSIK UND ASTRONOMIE
DIDAKTIK DER PHYSIK

Stefan Kirchner

Der Umgang mit Variablen beim Experimentieren
am Beispiel der Konstruktion
eines Wurfkörpers

Informierende Vorbereitung zum Experimentieren

KIRCHNER FOLIE 1

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Fragestellung

Wie muss ein Wurfkörper
konstruiert bzw. geformt werden,
damit er möglichst weit geworfen werden kann?

KIRCHNER FOLIE 2

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Beispiele möglicher Formen von Wurfkörpern, ...



... die unterschiedliche Aufgaben erfüllen !

KIRCHNER FOLIE 3

Fortsetzung A.1.9

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>Konkretisierung der Fragestellung ...</h3> <p>Wie muss man beim Experimentieren vorgehen, damit man einen Wurfkörper erhält, der möglichst weit geworfen werden kann?</p>	
KIRCHNER	FOLIE 4

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>... führt zum Thema</h3> <p>Der Umgang mit Variablen beim Experimentieren</p> <p style="text-align: center;">↓</p> <ul style="list-style-type: none"> • Welche Variablen haben einen Einfluss auf die Wurfweite ? (Variablen identifizieren & suchen, Planung einer Testreihe) • Wie ändert sich die Wurfweite bei system. Änderung einer Variablen ? (Verlauf & Effektstärken der Variablen innerhalb einer Testreihe, Bewertung) • Wie stark ist der Einfluss einer Variablen auf die Wurfweite ? (Gewichtung von Variablen, Vergleich & Bewertung von Testreihen) 	
KIRCHNER	FOLIE 5

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>Gliederung</h3> <ul style="list-style-type: none"> • <u>Einführung:</u> Begriffserklärungen, Variablen identifizieren, Ausgangsbeispiel • <u>Planung von Experimenten:</u> Experimentierplan mit Test- & Kontrollvariablen • <u>Umsetzung des Experimentierplans:</u> Experimentieranordnung & -demonstration zur Aufnahme der Messdaten • <u>Auswertung der realen Messdaten:</u> Vergleich & Bewertung von Variablen mittels Arbeitsblatt • <u>Abschluss:</u> Variablen suchen, Vertiefung, Zusammenfassung & Reflexion 	
KIRCHNER	FOLIE 6

Fortsetzung A.1.9

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>Einführung</h3> <p><u>Variablen:</u> Einflussgrößen, die beim Experimentieren verändert werden</p> <p><u>Unterscheidung der Variablen:</u></p> <ul style="list-style-type: none">• Variablen der Experimentierbedingungen, z.B. <i>Art & Spannkraft des Katapults, Standort des Wurfgerätes, ...</i>• Variablen des Wurfkörpers, z.B. <i>Masse, Durchmesser, Oberfläche, Material, ...</i> <p>➡ hier: Beschränkung auf Variablen des Wurfkörpers !!!</p>	
KIRCHNER	FOLIE 7

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>Einführung</h3> <p><u>Ausgangsbeispiel</u></p> <p>Wahl eines Wurfkörpers mit runder Grundform (Kugel) und den Ausgangsparametern</p> <p><i>Masse; Durchmesser; Oberfläche; Material</i> (15 g ; 10 cm ; glatt ; Styropor)</p>	
KIRCHNER	FOLIE 8

RUHR-UNIVERSITÄT BOCHUM	DIDAKTIK DER PHYSIK
<h3>Einführung</h3> <p>Abb. Ausgangsbeispiel</p>  <p><u>Mögliche Fragestellungen zur Untersuchung von Variablen</u></p> <ul style="list-style-type: none">• Hat die Kugelmasse einen Einfluss auf die Wurfweite ?• Hat der Durchmesser einen Einfluss auf die Wurfweite ?• Hat die Beschaffenheit der Oberfläche einen Einfluss auf die Wurfweite ?• ...	
KIRCHNER	FOLIE 9

Fortsetzung A.1.9

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Kugelmasse

The seal of the Ruhr-Universität Bochum is located in the top right corner. It is a circular emblem with the university's name 'RUHR-UNIVERSITÄT BOCHUM' around the perimeter. In the center, there is a stylized figure, possibly a scholar or a saint, holding a book or a staff.

	<u>Masse</u>	<u>Durchmesser</u>	<u>Oberfläche</u>	<u>Material</u>
Kugel	(15 g	10 cm	glatt	Styropor)
Kugel	(25 g	10 cm	glatt	Styropor)
Kugel	(35 g	10 cm	glatt	Styropor)
Kugel	(45 g	10 cm	glatt	Styropor)
Kugel	(55 g	10 cm	glatt	Styropor)
Kugel	(:	:	:	:

- Experimentierplan, der vergleichende Untersuchung zum Test der Variablen Masse ermöglicht

KIRCHNER

FOLIE 10

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Kugelmasse

Testvariable:

Variable, die verändert wird

	<u>Masse</u>	<u>Durchmesser</u>	<u>Oberfläche</u>	<u>Material</u>
Kugel	(15 g	10 cm	glatt	Styropor)
Kugel	(25 g	10 cm	glatt	Styropor)
Kugel	(35 g	10 cm	glatt	Styropor)
Kugel	(45 g	10 cm	glatt	Styropor)
Kugel	(55 g	10 cm	glatt	Styropor)
Kugel	(:	:	:	:)

KIRCHNER

FOLIE 11

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Kugelmasse

Kontrollvariablen:

Variablen, die nicht verändert werden

Masse

Durchmesser

Oberfläche

Material

Kugel	(15 g	10 cm	glatt	Styropor)
Kugel	(25 g	10 cm	glatt	Styropor)
Kugel	(35 g	10 cm	glatt	Styropor)
Kugel	(45 g	10 cm	glatt	Styropor)
Kugel	(55 g	10 cm	glatt	Styropor)
Kugel	(:	:	:	:

KIRCHNER

FOLIE 12

Fortsetzung A.1.9

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Kugelmasse

	<u>Masse</u>	<u>Durchmesser</u>	<u>Oberfläche</u>	<u>Material</u>
Kugel (15 g	10 cm	glatt	Styropor)	
Kugel (25 g	10 cm	glatt	Styropor)	
Kugel (35 g	10 cm	glatt	Styropor)	
Kugel (45 g	10 cm	glatt	Styropor)	
Kugel (55 g	10 cm	glatt	Styropor)	
Kugel (:	:	:	:	

Dimensionen der Variablen

KIRCHNER

FOLIE 13

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan zur Untersuchung der Variable Durchmesser

Testvariable

Masse Durchmesser Oberfläche Material

Kugel	(:	:	:	:)
Kugel	(15 g	8 cm	glatt	Styropor)
Kugel	(15 g	9 cm	glatt	Styropor)
Kugel	(15 g	10 cm	glatt	Styropor)
Kugel	(15 g	11 cm	glatt	Styropor)
Kugel	(15 g	12 cm	glatt	Styropor)
Kugel	(:	:	:	:)

Kontrollvariablen


KIRCHNER

FOLIE 14

RUHR-UNIVERSITÄT BOCHUM

DIDAKTIK DER PHYSIK

Experimentierplan



	<u>Masse</u>	<u>Durchmesser</u>	<u>Oberfläche</u>	<u>Material</u>
Kugel	(:)	:	:	:
Kugel	(15 g)	8 cm	glatt	Styropor
Kugel	(20 g)	9 cm	glatt	Styropor
Kugel	(25 g)	10 cm	glatt	Styropor
Kugel	(15 g)	11 cm	10 Löcher	Styropor
Kugel	(15 g)	12 cm	glatt	Gummi
Kugel	(:)	:	:	:



**Warum ist eine vergleichende Untersuchung
zum Test einer Variablen mit diesem
Experimentierplan nicht möglich?**

KIRCHNER

FOLIE 15

Fortsetzung A.1.9

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Abschluss: Zusammenfassung und Reflexion			
Variablen identifizieren, Variablen suchen	Var. 1; Var. 2; Var. 3; ... Bsp.: Breite; Winkel; Anzahl; ...		
Variablen testen, und kontrollieren	Testvar. 2 (Bsp. Winkel) Kontrollvar. 1; 3; ... (Bsp.: Breite; Anzahl; ...)		
Variablen vergleichen, Variablen bewerten	Gewichtung (Var 2 > Var 1 > Var 3 ...) Bsp. (Winkel > Breite > Anzahl ...) Effektstärke (Var. 1, 2 > Var. 3 ...) Bsp. (Breite, Winkel > Anzahl ...)		
KIRCHNER		FOLIE 16	

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Abschluss: Zusammenfassung und Reflexion			
Variablen identifizieren, Variablen suchen	 	ist eine Voraussetzung für ist eine Voraussetzung für	
Variablen testen, und kontrollieren			
Variablen vergleichen, Variablen bewerten			
KIRCHNER		FOLIE 17	

RUHR-UNIVERSITÄT BOCHUM		DIDAKTIK DER PHYSIK	
Vielen Dank für Eure Aufmerksamkeit			
KIRCHNER		FOLIE 18	

A.1.10

Messung maximaler Wurfweiten von Wurfkörpern**1. Ausgangsparameter eines Wurfkörpers mit kugeliger Grundform:****Masse (14,0 g), Durchmesser (10,0 cm), Oberfläche (glatt), Material (Styropor)****2. Messfolgen zum Test der Variablen****a) Masse, b) Durchmesser und c) Oberfläche (Anzahl Löcher) mit der Federhärte des Katapults $D = 2,2 \text{ N/cm}$**

2 a)	
<i>Masse</i> in g	max. Wurfweite in cm
⋮	⋮
15,0	840
25,0	822
35,0	756
45,0	690
55,0	624
⋮	⋮
<u>Testvariable:</u> <i>Masse</i>	<u>Kontrollvariablen:</u> Durchmesser (10 cm) Oberfläche (glatt) Material (Styropor)

2 b)	
<i>Durchmesser</i> in cm	max. Wurfweite in cm
⋮	⋮
4,0	1160
5,0	1163
6,0	1157
7,0	1125
8,0	958
⋮	⋮
<u>Testvariable:</u> <i>Durchmesser</i>	<u>Kontrollvariablen:</u> Masse (14,0 g) Oberfläche (glatt) Material (Styropor)

2 c)	
<i>Anzahl Löcher</i>	max. Wurfweite in cm
0 (glatt)	780
10	805
20	814
30	804
40	786
⋮	⋮
<u>Testvariable:</u> <i>Oberfläche</i>	<u>Kontrollvariablen:</u> Masse (14,0 g) Durchmesser (10,0 cm) Material (Styropor)

Fortsetzung A.1.10

3. Graphische Auswertung der Messfolgen

(Federhärte des Katapults $D = 2,2 \text{ N/cm}$)

Diagramm 1

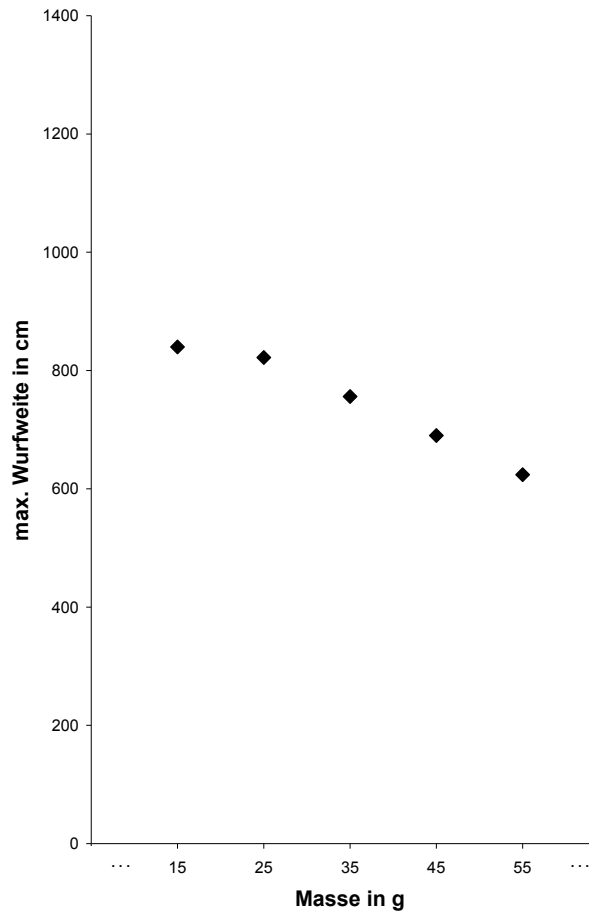


Diagramm 2

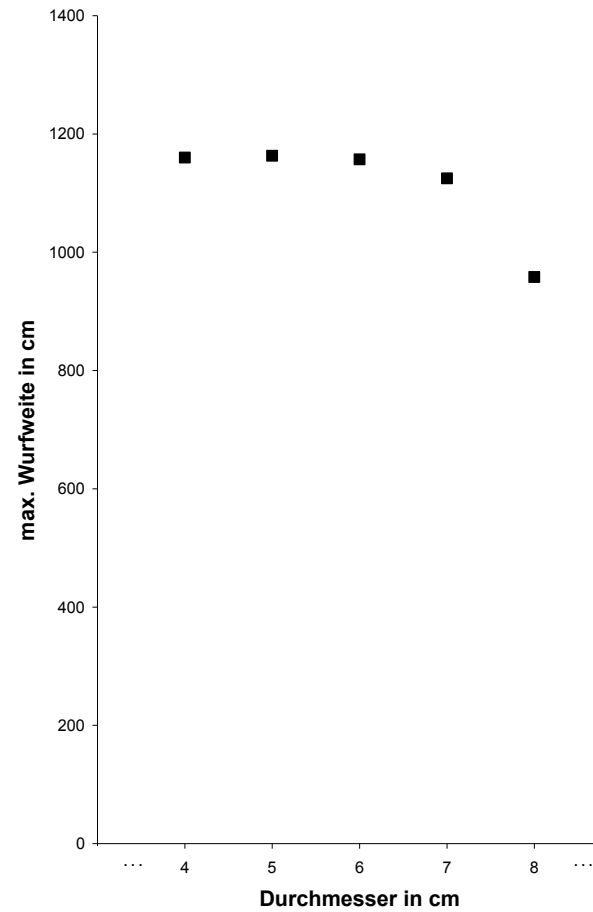


Diagramm 3

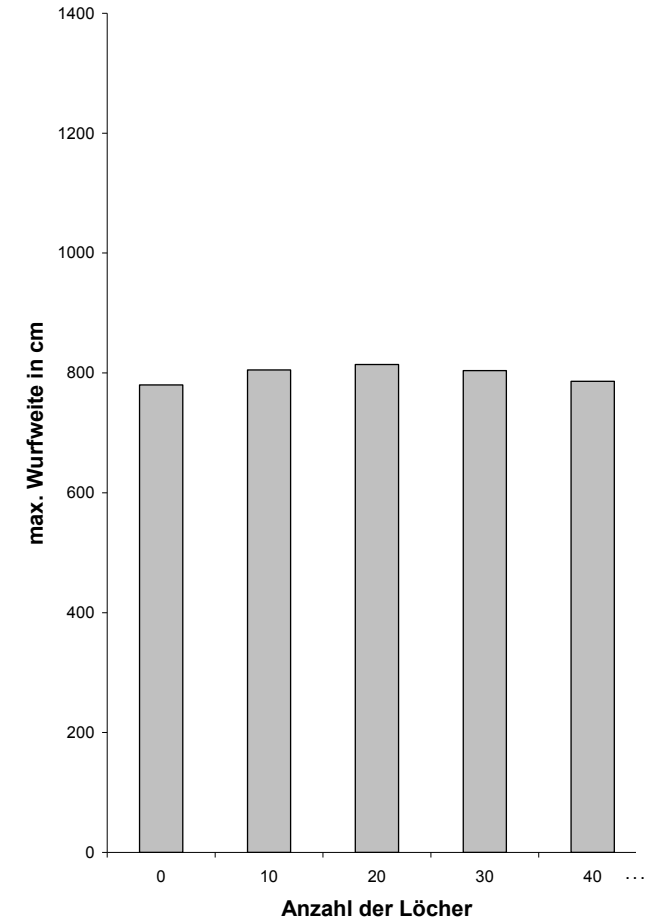


Diagramme: Abhängigkeit der maximalen Wurfweite von der Masse (1), vom Durchmesser (2) und von der Anzahl der Löcher in der Oberfläche (3) des Wurfkörpers

A.2.1

Messinstrument Fach- & NAW-Test Aufbau – Struktur – Bewertung						
Nr.	Item	Quelle	Dimension		Punkte	
1 2 3 4	Aufgabeneinheit: Windenergie					
	Frage 1	PI	–	NAW	–	1
	Frage 2	PI	–	NAW	–	1
	Frage 3	PI	Fach	–	1	–
	Frage 4	PI	Fach	–	1	–
5 6 7 8	Aufgabeneinheit: Statischer Auftrieb (SA)					
	Frage 1	EK	–	NAW	–	1
	Frage 2	EK	–	NAW	–	1
	Frage 3	EK	–	NAW	–	1
	Frage 4	EK	–	NAW	–	1
9 10 11 12	Aufgabeneinheit: Sonnenschutz					
	Frage 1	PI	–	NAW	–	1
	Frage 2	PI	–	NAW	–	1
	Frage 3	PI	–	NAW	–	1
	Frage 4	PI	–	NAW	–	1
13	O 13	TI 2	Fach	–	1	–
14	P 7	TI 2	–	NAW	–	1
15	I 12	TI 2	–	NAW	–	1
16	DA	EK	Fach	–	2	–
17	L 1	TI 2	Fach	–	1	–
18	N 1	TI 2	–	NAW	–	1
19	B 6	TI 3	Fach	–	1	–
20	K 17	TI 2	Fach	–	1	–
21	H 4	TI 3	Fach	–	1	–
22	G 11	TI 3	Fach	–	2	–
23	Y 1	TI 2/3	Fach	–	1	–
24	G 15	TI 3	Fach	–	1	–
Σ	24	–	11	13	13	13
TI 2/ 3 PI, EK DA, Y 1	Anmerkung: TIMS-Studie Population 2 (TI 2) und Population 3 (TI 3) PISA-Studie 2006 (PI), Eigenkonstruktion (EK) Dynamischer Auftrieb (DA), Anker-Item TIMS-Studie (Y 1)					

Tab. A 4: Messinstrument Fach- & NAW-Test. Aufbau – Struktur – Bewertung

Fortsetzung A.2.1**Fragen zu Deiner Person****Datum:** _____

1. Damit Fragebögen zu unterschiedlichen Zeitpunkten einander zugeordnet werden können, man aber nicht weiß, wer genau den Fragebogen ausgefüllt hat, wird jeder befragten Person ein Erkennungscode zugeordnet. Fülle diesen bitte zunächst aus.

ERKENNUNGSCODE

Die ersten beiden Buchstaben des Vornamens deiner Mutter.		Die ersten beiden Buchstaben des Vornamens deines Vaters.	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

2. Ich bin

☐ männlich ☐ weiblich

3. Mein Geburtsdatum: _____

4. Welche Note in Physik hattest du auf dem Abschlusszeugnis der Klasse 10?

1	2	3	4	5	6	kein Physikunterricht
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. Für die Jahrgangsstufe 11 habe ich folgende Grundkurse gewählt:

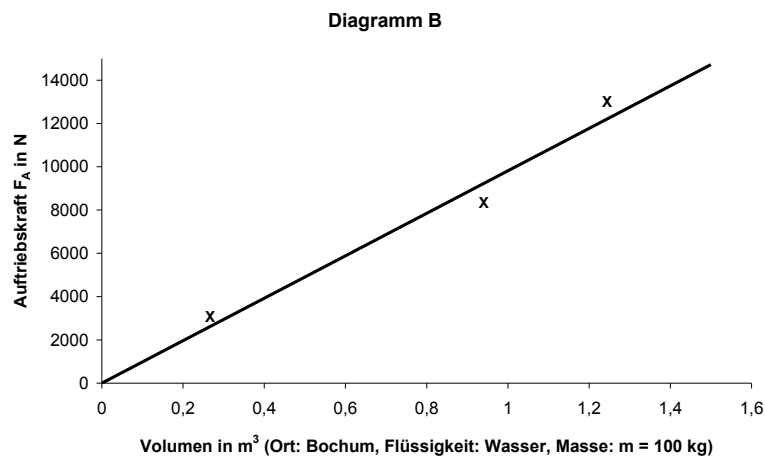
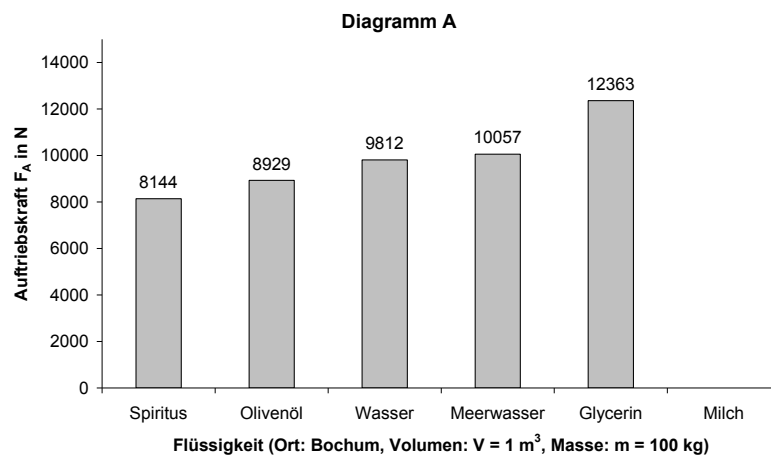
Physik	Chemie	Biologie
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fortsetzung A.2.1**Aufgabeneinheit SA: Statischer Auftrieb in Flüssigkeiten**

Körper innerhalb von Flüssigkeiten erfahren Auftriebskräfte, wie sie beispielsweise beim Tauchen auftreten. Die Größe der Auftriebskraft F_A hängt ab von:

- (1) dem Volumen V des untergetauchten Körpers,
- (2) der Dichte ρ der Flüssigkeit,
- (3) der Erdbeschleunigung g (auch Ortsfaktor g genannt) an diesem Ort

nach der Gleichung: $F_A = \rho \cdot V \cdot g$.

Diagramme A bis D zu: Statischer Auftrieb in Flüssigkeiten

Fortsetzung Diagramme zur Aufgabeneinheit SA

Diagramm C

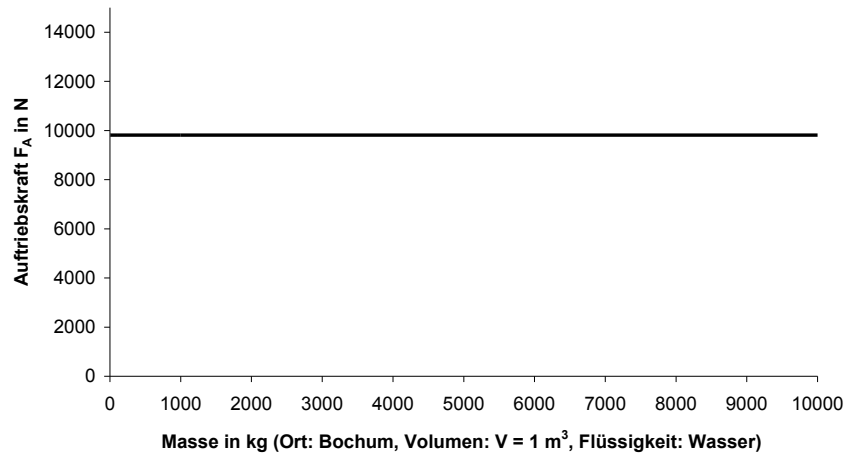
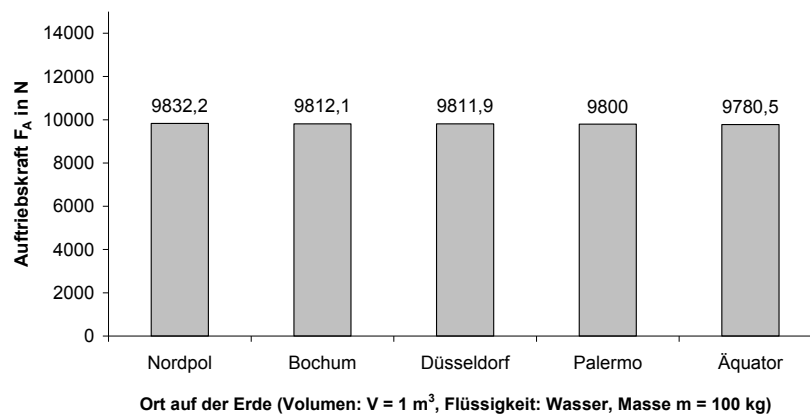


Diagramm D



Anzahl der Fragen zur Aufgabeneinheit SA: 4

Fragen zur Aufgabeneinheit SAFrage 1:

Ein Körper wird vollständig in verschiedene Flüssigkeiten (siehe Diagramm A) mit folgenden Dichten ρ getaucht: $\rho_{\text{Spiritus}} = 830 \text{ kg/m}^3$, $\rho_{\text{Olivenöl}} = 910 \text{ kg/m}^3$, $\rho_{\text{Wasser}} = 1000 \text{ kg/m}^3$, $\rho_{\text{Meerwasser}} = 1025 \text{ kg/m}^3$, $\rho_{\text{Glycerin}} = 1260 \text{ kg/m}^3$. Das Diagramm A stellt die jeweilige Auftriebskraft F_A dar.

Skizziere (ohne Rechnung) in dem Diagramm A, welche Auftriebskraft der Körper in Milch mit der Dichte $\rho_{\text{Milch}} = 1030 \text{ kg/m}^3$ erfahren würde.

Frage 2:

Die Diagramme A bis D stellen die Auftriebskraft F_A eines Körpers in Abhängigkeit folgender Variablen dar:

- (a) ausgewählte Flüssigkeiten, in die er getaucht wird,
- (b) seines Volumens,
- (c) seiner Masse,
- (d) des Ortes, an dem die Messung erfolgt.

Ordne mit den Angaben im Diagramm die Variablen (a) bis (d) nach der Stärke ihres Einflusses auf die Auftriebskraft F_A , beginnend mit dem größten Einfluss.

Ordnung der Variablen:.....

Lösung: VAR (Volumen), VAR (Dichte), VAR(Ort), VAR (Masse)

Frage 3:

Ein Würfel wird vollständig in Wasser getaucht. Welche der folgenden Aussagen ist richtig?

- (a) Je größer die Oberfläche des Würfels, desto kleiner ist die Auftriebskraft F_A .
- (b) Je größer die Oberfläche des Würfels, desto größer ist die Auftriebskraft F_A .
- (c) Die Größe der Oberfläche hat keinen Einfluss auf die Auftriebskraft F_A .
- (d) Die Größe der Oberfläche hat nur einen ganz geringen Einfluss auf die Auftriebskraft F_A .
- (e) Man kann keine Aussage über den Einfluss der Oberfläche auf die Auftriebskraft F_A machen.

Lösung: Antwort (b)

Fortsetzung Fragen zur Aufgabeneinheit SA

Frage 4:

Im Diagramm B wurde aufgrund von drei Messwerten (kleine Kreuze) eine Gerade eingezeichnet. Welche Aussage über den graphischen Zusammenhang zwischen Volumen V und Auftriebskraft F_A ist nur aufgrund der vorliegenden Messungen (kleine Kreuze) berechtigt?

- (a) Man kann anhand der drei Messungen direkt auf eine Gerade schließen. (Es reichen schon zwei Punkte, um eine Gerade festzulegen.)
- (b) Man kann anhand der drei Messungen eine Gerade ausschließen. (Die drei Punkte liegen nicht auf einer Geraden.)
- (c) Man kann anhand der drei Messungen auf eine andere Funktion schließen. (Die drei Messwerte liegen eher auf einer anderen Funktion als auf einer Geraden.)
- (d) Man kann anhand der drei Messungen eine andere Funktion ausschließen. (Geraden kommen in der Physik oft vor, die Abweichungen stammen von Messfehlern.)
- (e) Man kann anhand der drei Messungen keine Aussagen treffen. (Man braucht mehr Messwerte, um eine Aussage treffen zu können.)

Lösung: Antwort (e)

Testitem DA: Dynamischer Auftrieb

DA. Erkläre kurz, wie bei der Umströmung einer Tragfläche eines Flugzeuges die Auftriebskraft entsteht.

Testitem DA: Bewertung und Punktierung

Bewertung und Punktierung: Dynamischer Auftrieb		
Teillösung	Bewertung	Punkte
1	Anfahrwirbel an Hinterkante (Text oder Skizze); Zirkulationswirbel (tragender Wirbel) ist Folge des Anfahrwirbels (kein Wirbel ohne Gegenwirbel)	0,5
2	Überlagerung der Zirkulationsströmung mit der Luftanströmung (Text oder Skizze); Geschwindigkeitsdifferenz zwischen der Ober- & Unterseite der Tragfläche	0,5
3	Geschwindigkeit der Luftströmung ist auf der Oberseite der Tragfläche größer als auf der Unterseite der Tragfläche; Druckdifferenz zwischen der Ober- & Unterseite der Tragfläche	0,5
4	Unterdruck herrscht an der Tragflächenoberseite; Überdruck herrscht an der Tragflächenunterseite; Auftriebskraft ist nach oben gerichtet	0,5

A.2.2 BEOBACHTUNGSBOGEN**1. Daten des Probandenpaares****Datum der Beobachtung:** _____**ERKENNUNGSCODE Proband 1**

Die ersten beiden Buchstaben des Vornamens deiner Mutter.		Die ersten beiden Buchstaben des Vornamens deines Vaters.	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

ERKENNUNGSCODE Proband 2

Die ersten beiden Buchstaben des Vornamens deiner Mutter.		Die ersten beiden Buchstaben des Vornamens deines Vaters.	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

b) Geschlecht: ☐ m ☐ w☐ m ☐ w

c) Geburtsdatum: _____

a) Gruppenname für Wettbewerb: _____

b) Spannung U beim Wettbewerb: _____

c) Anzahl der leuchtenden LED's: _____

2. Daten der Schule

a) Name des Gymnasiums: _____

b) Name des Kurslehrers: _____

c) Vorbereitende Instruktion am: _____

zum Thema:

☐ Wind☐ Fliegen☐ Wurfweite**3. Daten des Beobachters**

a) Name des Beobachters: _____

b) Beginn der Beobachtung: _____

c) Ende der Beobachtung: _____

d) Videomitschnitt vorhanden (j/ n): _____

e) Dateiname des Videos: _____

f) schriftliches Material wie Tabellen, Skizzen, ... anbei? ☐ ja ☐ nein

Fortsetzung A.2.2

1. Informationen beschaffen & Profilkonstruktion planen

Seite 2

Zeit in min nach Beginn der Beobachtung		5	10	[...]	145	150
Informationen beschaffen: Informationsquelle						
	Infomappen			[...]		
	Internet					
	Nachbargruppe					
	Material der Vorbereitung (Arbeitsblatt, ...)					
	Setflügel					
Profilkonstruktion planen						
<i>Ein Profilentwurf wird</i>				[...]		
	... beschrieben (z. B. Diskussion)					
	... auf Papier skizziert					
<i>Detaillierte Abmessungen des Profils werden</i>						
	... benannt (z. B. nach Diskussion)					
	... auf Papier gezeichnet					
	... von Setflügel/ n übernommen					
<i>Mehrere Variablen (z. B. Länge, Breite,...) werden</i>						
	... genannt und/ oder notiert					
	... zu einem Profilentwurf kombiniert					
<i>Ein weiterer Faktor wird</i>						
	... genannt und/ oder notiert					
	... mit einem bestehenden Profilentwurf kombiniert					
<i>Zu testende Variablen werden</i>						
	... benannt					
	... notiert (z. B. Messtabelle anlegen)					
Anmerkungen:						

Fortsetzung A.2.2

2. Arbeiten mit den Flügelsets

Seite 3

Zeit in min nach Beginn der Beobachtung		5	10	[...]	145	150
Transport von Flügeln zum Arbeitsplatz						
	vollständiges Set: Set-Nr./ n			[...]		
	Anzahl einzelner Flügel der Set-Nr.					
Arbeiten mit Flügeln der Set-Nr.						
Flügel am Generator montieren:						
	alle Flügel des Sets montieren			[...]		
	Anzahl der gewählten & montierten Flügel					
Flügel testen						
Flügel anströmen						
	alle montierten Flügel anströmen			[...]		
	Anzahl der montierten & angeströmten Flügel					
	Flügel umgekehrt montieren & anströmen					
Messdaten aufnehmen						
	Messung mit Voltmeter durchführen			[...]		
	Messdaten nicht beachten (ignorieren)					
	Messdaten ablesen					
	alle Messdaten schriftlich notieren					
	Anzahl der schriftlich notierten Messdaten					
	Messung wiederholen (weiter Seite 5)					
Kontrollbedingungen des Versuchs						
	werden beim Ablesen/ Notieren nicht eingehalten			[...]		
	werden verändert (z. B. explorieren, ...)					
Experimenteller Erfolg-Misserfolg						
	Windrad dreht sich nicht			[...]		
	Windrad dreht sich anders als erwartet (besser, ...)					
Experimentelle Fehler und Fehlersuche						
	falsche Montage (lockere, schleifende, ... Flügel)			[...]		
	Messdaten fehlerhaft ablesen (Messbereich, ...)					
	kein Zeigerausschlag					
	rechts-/ linksdrehende Flügel, Anstellwinkel, ...					
	Fehler erfolgreich beheben					
Anmerkungen:						

Fortsetzung A.2.2

3. Arbeiten mit Flügelsetkombinationen

Seite 4

Zeit in min nach Beginn der Beobachtung		5	10	[...]	145	150
Nummerierung der Flügelsets:						
Identische Flügel am Generator montieren:						
	mehrere identische Flügel der Set-Nr. montieren			[...]		
	Anzahl identischer Flügel verändern					
Flügelkombinationen am Generator montieren						
	Flügel der Set-Nr./ n montieren			[...]		
	Anzahl der Flügelkombinationen <u>eines</u> Sets					
	Anzahl der Flügelkombinationen <u>diverser</u> Sets					
Flügel testen						
Flügel anströmen						
	alle montierten Flügel anströmen			[...]		
	Anzahl der montierten & angeströmten Flügel					
	Flügel umgekehrt montieren & anströmen					
Messdaten aufnehmen						
	Messdaten mit Voltmeter durchführen			[...]		
	Messdaten nicht beachten (ignorieren)					
	Messdaten ablesen					
	Messdaten schriftlich notieren (Anzahl, alle, ...)					
	Messung wiederholen (weiter Seite 5)					
Kontrollbedingungen des Versuchs						
	werden beim Ablesen/ Notieren nicht eingehalten			[...]		
	werden bewusst verändert (z. B. explorieren, ...)					
Experimenteller Erfolg-Misserfolg						
	Windrad dreht sich nicht			[...]		
	Windrad dreht sich anders als erwartet (besser, ...)					
Experimentelle Fehler und Fehlersuche						
	falsche Montage (lockere, schleifende, ... Flügel)			[...]		
	Messdaten fehlerhaft ablesen (Messbereich, ...)					
	kein Zeigerausschlag					
	rechts-/ linksdrehende Flügel, Anstellwinkel, ...					
	Fehler erfolgreich beheben					
Anmerkungen:						

Fortsetzung A.2.2

4. Messwiederholungen

Seite 5

Zeit in min nach Beginn der Beobachtung		5	10	[...]	145	150
Nummerierung der Flügelsets:						
Messwiederholungen & Vergleich von ...						
	Messergebnisse der Vorbereitung "Wind" prüfen			[...]		
	vollständiges Set wiederholt messen: Set-Nr.					
	Anzahl einzelner Flügel der Set-Nr.					
	Anzahl Flügelkombinationen (identische Flügel)					
	Anzahl Flügelkombinationen (nicht ident. Flügel)					
	nur Maximalwerte auswählen					
	Mittelwert aus Messwiederholung bilden					
	Messergebnisse wiederholt notieren					
	Messwiederholung ergibt eine Abweichung					
	Messwiederholung ergibt keine Abweichung					
<i>Abweichung</i>						
	... feststellen					
	... als gering einstufen					
	... ignorieren					
	... mit veränderten Kontrollbedingungen erklären					
	... mit veränderter Montage erklären					
	... nicht erklärbar					
	...					
Anmerkungen:						

Fortsetzung A.2.2

5a. Messdaten vergleichen und bewerten

Seite 6

Zeit in min nach Beginn der Beobachtung	5	10	[...]	145	150
Nummerierung der Flügelsets:					
Vergleich der Messdaten von <u>Einzelmessungen</u>					
Messdaten ignorieren & nicht vergleichen			[...]		
alle Einzelmessungen vergleichen					
Einzelmessungen nur zum Teil vergleichen					
Maximalwerte aus Einzelmessungen bestimmen					
Vergleich der Messdaten <u>eines</u> Flügelsets: Set-Nr.					
<i>Datenbasis zum Vergleich</i>					
Messdaten mit Vorbereitung "Wind" vergleichen			[...]		
alle Messdaten eines Flügelsets vergleichen					
nicht alle Messdaten eines Flügelsets vergleichen					
Maximalwert eines Flügelsets bestimmen					
<i>Erweiterung der Datenbasis zum Vergleich</i>					
Fehlen von Messdaten feststellen			[...]		
Fehlen von Zwischenwerten feststellen					
Intervallgrenzen der Flügelsets feststellen					
Gültigkeit der Daten auf Messintervall begrenzen					
möglichen Verlauf über Intervallgrenze vermuten					
Messverlauf über Intervallgrenze postulieren					
Feststellung: Sets zum Test "neuer" Variablen fehlen					
<i>Rangfolge & Effektstärken der Messdaten</i>					
Effektstärke diskutieren, feststellen			[...]		
Effektstärke auf Messintervall begrenzen					
Rangfolge der Maximalwerte erstellen					
Rangfolge mit Daten der Vorbereitung erstellen					
<i>Verlauf einer Variablen feststellen</i>					
keine Änderung im Verlauf feststellen			[...]		
Extrema von Messdaten feststellen					
Linearität von Messdaten feststellen					
Verlauf auf Messintervall begrenzen					
Anmerkungen:					

Fortsetzung A.2.2

5b. Fortsetzung Messdaten vergleichen und bewerten

Seite 7

Zeit in min nach Beginn der Beobachtung		5	10	[...]	145	150
Nummerierung der Flügelsets:						
Vergleich der Messdaten <u>mehrerer</u> Flügelsets						
	Daten mit Daten der Vorbereitung vergleichen			[...]		
	Messdaten aller Flügelsets vergleichen					
	Messdaten nicht aller Flügelsets vergleichen					
	nur Maximalwerte der Flügelsets vergleichen					
	Unterschiede in Variablenverläufen feststellen					
	Rangfolge der Maximalwerte erstellen					
	Effektstärken der Flügelsets vergleichen					
	...					
Vergleich der Messdaten <u>kombinierter</u> Flügelsets						
	Messdaten nicht vergleichen			[...]		
	Kombinationen eines Sets vergleichen					
	Kombinationen mehrerer Sets vergleichen					
	Kombinationen mit Vorbereitung vergleichen					
	Kombinationen mit Flügelset vergleichen					
	...					
Bewertung von Messergebnissen						
	<i>Beim Vergleich von Messergebnissen werden</i>			[...]		
	... Begründungen über Verläufe gegeben					
	... Begründungen von Einflüssen gegeben					
	... je-desto-Beziehungen aufgestellt					
	... Plausibilitätserklärungen gegeben					
	... abhängige & unabhängige Variablen diskutiert					
	... Sets mit anderen Ausgangsparametern bewertet					
	...					
Anmerkungen:						

Fortsetzung A.2.2

Protokoll der Profilkonstruktionen

Seite 8 ... N

Zeit [5 min]	Protokoll der Beobachtungen	Profilskizze

Fortsetzung A.2.2

Beiblatt Seite 1 Spalte 1

1 Planung von Profilen
<p><i>Planungen vor der Konstruktion</i></p> <p>Schüler haben eine Idee ...</p> <p>Schüler planen ein Profil ...</p> <p>...</p> <p><i>Ein Profilentwurf wird</i></p> <p>... beschrieben (z. B. Diskussion)</p> <p>... auf Papier skizziert</p> <p>...</p> <p><i>Detaillierte Profilmessungen werden</i></p> <p>... benannt (z. B. nach Diskussion)</p> <p>... auf Papier gezeichnet</p> <p>... von Setflügel/ n übernommen</p> <p>...</p> <p><i>Mehrere Variablen (Länge, Breite, ...) werden</i></p> <p>... genannt und/ oder notiert</p> <p>... zu einem Profilentwurf kombiniert</p> <p>...</p> <p><i>Ein weiterer Faktor wird</i></p> <p>... genannt und/ oder notiert</p> <p>... mit einem bestehenden Profilentwurf kombiniert</p> <p>...</p> <p><i>Systematisch zu testende Variablen werden</i></p> <p>... benannt</p> <p>... notiert (z. B. Messtabelle anlegen)</p> <p>...</p> <p><i>Ideen & Ziele der Konstruktion</i></p> <p><i>Ziele der Konstruktion sind</i></p> <p>... Maximalwerte aller Variablen zu vereinen</p> <p>... nur die größten Maximalwerte zu vereinen</p> <p>... Variablen mit großer Effektstärke zu vereinen</p> <p>... ein Profil aus den Flügelsets zu kopieren</p> <p>... ein funktionierendes Windrad zu bauen</p> <p>... die Konstruktion schnell zu beenden</p> <p>...</p>

Fortsetzung A.2.2

Beiblatt Seite 1 Spalte 2

2a Konstruktion von Profilen

Profilkonstruktion ist

- ... erste Konstruktion (Profil 1)
- ... ein Neubeginn (Profil 2, 3, ...)

Konstruktion erfolgt/ Profilform entsteht ...

- ... auf Basis der experimentellen Arbeit mit den Sets
- ... durch Kopie eines Setflügels
- ... durch Vereinigung von Maximalwerten, ...
- ... ad hoc, nach schönem Aussehen
- ... ohne experimentelle Vorinformation
- ... nur auf Basis skizzierter Flügel
- ... nach technischen Vorbildern, z. B. Windkraftanlagen

Allgemeines zur Profilkonstruktion

Konstruktion eines Profils wird

- ...nach Skizze anfertigt
- ... nach detaillierter Vorlage gebaut
- ... mit konkreten Maßen auf PVC gezeichnet
- ... ohne zuvor erstellte Vorlage angefertigt
- ... beendet
- ... nicht fertig gestellt
- ... fortgesetzt (z. B. Profil 3, ...)

Planungsänderung und -ergänzung

Während der Konstruktion wird/ werden

- ... Abmessungen detailliert festgelegt
- ... Abmessungen der Vorlagen verändert
- ... eine neue Variable genannt
- ... Profil um neue Variable ergänzt
- ... zu testende Variable genannt
- ... Verträglichkeit mehrerer Maximalwerte diskutiert
- ... Vereinigung mehrerer Maximalwerte diskutiert

...

Hinweis für den Beobachter

Grundformen der Profile ...

- ... bitte skizzieren !
- ... kurz beschreiben !
- ... nummerieren (Profil 1, 2, ...) !

Fortsetzung A.2.2

Beiblatt Seite 1 Spalte 3

2b Konstruktion von Profilen***Konstruktion auf Basis der Arbeit mit Sets****Konstruktion erfolgt unter Einschluss*

... der Variablen mit maximaler Gewichtung

... der Variablen mit maximaler Effektstärke

... von Variablen mit großer Gewichtung

... von Variablen mit großen Effektstärken

Konstruktion erfolgt durch Vereinigung

... der beiden größten Maximalwerte

... weiterer Maximalwerte

... der Variablen mit großer Effektstärke

... aller Maximalwerte der getesteten Variablen

Konstruktion erfolgt unter Ausschluss

... von Variablen, z. B. "ohne Einfluss",

... von Variablen, z. B. geringe Effektstärken

... von Variablen mit geringer Gewichtung

Konstruktion durch Kopie von Flügelsets*Kopieren eines Setflügels, Kopieren*

... eines Flügels (Nr.)

... mehrerer, identischer Flügel (Nr., Anzahl)

... mit Maximalwerten

... mit maximaler Effektstärke

... mit Vereinigung von Maxima weiterer Variablen

... unter Ausschluss von Variablen, z. B. "ohne Einfluss"

... mit Änderungen zum Original (neue Länge, ...)

... eines willkürlich ausgewählten Setflügels

Kopieren einer Flügelkombination, Kopieren ...

... einer Kombination mit maximaler Gewichtung, ...

... einer Kombination mit maximaler Effektstärke, ...

... einer willkürlich gewählten Kombination

...

Hinweis für den Beobachter***Bitte protokollieren !***

Welche Variablen werden ...

Wie häufig werden Variablen ...

Welche Dimensionen der Variablen werden ...

vereinigt, kopiert, eingeschlossen, ...

Fortsetzung A.2.2

Beiblatt Seite 1 Spalte 4

3 Änderungen an bestehenden Profilkonstruktionen*Allgemeines*

Veränderungen erfolgen

... am Profil Nr. 1, 2, 3, usw.

... systematisch (Messfolge)

... unkontrolliert usw.

... schon vor Ende einer Konstruktion

... ohne ein Profil getestet zu haben

...

*Änderungen einer Variablen**Eine Variable wird*

... einmal verändert

... mehrfach verändert (Anzahl der Änderungen)

... verändert im Intervall (5°-30°, 0-50 Löcher, ...)

... verändert mit den Intervallschritten (10°-Schritte, ...)

Welche Variable wird verändert?

Verdrehung, Winkel, Länge, Winglet, Breite, ...

Oberfläche, Material, ...

Änderung der Variablen erfolgt... an einem Profil (verschiedene Schliffstärken, ...)

... durch Dublizieren von Profilen (Anzahl)

... mit Kontrolle der Ausprägung (Winkel einstellen, ...)

... nach Augenmaß

...

*Unkontrollierte Änderung von Variablen**Profil wird unkontrolliert verändert*

... durch willkürliches Biegen, Kürzen, usw.

... durch fortwährendes Schleifen, Biegen usw.

... ohne Kontrolle der Ausprägung/ des Augenmaßes

...

Während einer Änderung wird/ werden

... mehr als eine Variable verändert

... mehrere Änderungen gleichzeitig vorgenommen

... mehrere Materialien kombiniert (PVC, ...)

... zwei unterschiedliche Profilformen kombiniert

... unterschiedliche Profilformen systematisch getestet

Fortsetzung A.2.2

Beiblatt Seite 2 Spalte 1

4 Messfolge erstellen

Ziel der Messfolge ist

- ... eine (neue) Variable an eigenen Profilkonstruktionen zu testen
- ... ein bestehendes Flügelset zu erweitern (Zwischenwert, Intervallgrenze)
- ... ein bestehende Flügelset mit neuen Ausgangsparametern zu testen
- ... die "Verträglichkeit" zweier Variablen mit großer Effektstärke zu testen
- ... die "Abhängigkeit" von Variablen aufzulösen (Länge/ Masse, ...)
- ... ein systematischer Test von Profilkombinationen

...

"Neue Variablen" (Masse, Massenverteilung, Oberfläche, ...)

Test neuer Variablen durch

- ... Materialwechsel: Metall, Holz, ...
- ... Materialwechsel: dickes, dünnes PVC
- ... Oberflächenänderung: durchlöchern, aufrauen, schleifen, ...
- ... Änderung der Profilenden bzw. -kanten: spitz, breit, ...

...

Ausprägung der Messfolge

Die Testvariable wird

- ... einfach, mehrfach verändert (Anzahl der Änderungen)
- ... verändert im Intervall (5° - 30°, 0 - 50 Löcher, ...)
- ... verändert mit den Intervallschritten (10°-Schritte, viel Schliff, ...)

Änderung der Testvariablen erfolgt

- ... an einem Profil (verschiedene Schliffstärken, ...)
- ... durch Konstruktion eines neuen Flügelsets (Anzahl der Flügel)
- ... mit Kontrolle der Ausprägung (Winkel einstellen, ...)
- ... nach Augenmaß (Winkel, Längen, ...)

...

Fortsetzung A.2.2

Beiblatt Seite 2 Spalte 2

5 Testung von Profilen***Beim Test****Umsetzung Konstruktionsplan*

- ... handelt es sich um ein Zwischentest
- ... ist Plan nicht, zum Teil oder vollständig umgesetzt
- ... ist ursprünglicher Plan bereits verworfen

Experimenteller Erfolg-Misserfolg

- ... dreht sich das Windrad nicht
- ... besser, schlechter, anders als erwartet
- ... ist die Spannung U

Messdaten aufnehmen

- ... erfolgt Messung mit Voltmeter/ LED
- ... werden Messdaten nicht beachtet (ignoriert)
- ... werden Messdaten abgelesen/ notiert

Kontrollbedingungen des Versuchs

- ... werden beim Ablesen/ Notieren nicht eingehalten
- ... werden (bewusst) verändert

Experimentelle Fehler und Fehlersuche

- ... wird Profil fehlerhaft montiert (lockere Flügel, ...)
- ... werden Daten falsch abgelesen
- ... ist kein Zeigerausschlag vorhanden
- ... hat Profil system. Fehler (kein Anstellwinkel, ...)
- ... wird Polung am Spannungsmesser umgekehrt
- ... wird Fehler erfolgreich behoben

Messung wiederholen

- ... werden Messungen wiederholt/ Anzahl Messwiederholungen
- ... werden Messwiederholungen verglichen
- ... wird Abweichung festgestellt, notiert, ...
- ... wird Abweichung als gering eingestuft, ignoriert, erklärt, ...
- ... wird ein/ werden alle Maximalwert(e) berücksichtigt
- ... wird Mittelwert aus Messwiederholungen gebildet

...

Nach einem Zwischentest/ Test*Konstruktionsplan*

- ... wird Plan fortgesetzt, verändert, verworfen, ...
- ... wird ein anderer, neuer, ... Plan verfolgt
- ... findet ein Planungswechsel/ eine Umorientierung, statt

Fortsetzung A.2.2

Beiblatt Seite 2 Spalte 3

6 Messergebnisse vergleichen und bewerten***Allgemeines zum Vergleich****Messdaten der Konstruktion werden verglichen*

- ... mit Daten eines Flügels der Set-Nr.
- ... mit Daten eines/ mehrerer Flügelsets (Nr.)
- ... mit Daten einer/ mehrerer Flügelsetkombinationen
- ... mit Daten einer neuen Variablen
- ... mit Profilkonstruktion-Nr. 2, 4, 3, ...
- ... mit Messwiederholungen desselben Profils
- ... mit Daten vor der Profiländerung
- ... mit Einzelmessungen, Daten von Nachbargruppen, usw.

Datenbasis zum Vergleich*Von den getesteten Profilen wird/ werden ...*

- ... Daten nicht verglichen oder ignoriert
- ... Daten nur zum Teil/ alle Daten verglichen
- ... nur Maximalwerte verglichen
- ... Gültigkeit der Daten auf Intervallgrenze begrenzt
- ... Messverläufe über Intervallgrenzen vermutet

Rangfolge und Effektstärken von Messdaten*Beim Vergleich von Daten wird Rangfolge*

- ... mit Daten der Vorbereitung "Wind" erstellt
- ... von Einzelmessungen erstellt

Beim Vergleich von Daten einer Messfolge

- ... werden Effektstärken diskutiert, festgelegt, ...
- ... werden Effektstärken auf Intervalle begrenzt

Beim Vergleich von Daten mehrerer Messfolgen

- ... werden Variablen/ Messdaten gewichtet

Verläufe von Variablen*Beim Vergleich von Messdaten, z. B. einer Messfolge wird*

- ... ein linearer/ ein extremer Variablenverlauf festgestellt
- ... keine Änderung im Variablenverlauf festgestellt
- ... der Variablenverlauf im begrenzten Intervall festgestellt

Bewertungen von Messergebnissen*Beim Vergleich von Messdaten werden*

- ... Begründungen über Variablenverläufe gegeben
- ... je-desto-Beziehungen aufgestellt
- ... Plausibilitätserklärungen gegeben
- ... abhängige und unabhängige Variablen diskutiert

A.3.1

Verteilung der Ergebnisse des NAW-Tests zum Messzeitpunkt 1 der Probandenpaare der Experimentalgruppen

	Kontrollgruppe				TG 1 (Reproduktion)				TG 2 (Reorganisation)				TG 3 (Transfer)			
P-Paar Nr.:	MZP 1			Statistik	MZP 1			Statistik	MZP 1			Statistik	MZP 1			Statistik
	Bereich / Punkte				Bereich / Punkte				Bereich / Punkte				Bereich / Punkte			
	P1	P2			P1	P2			P1	P2			P1	P2		
I	oben	13	13	Σ 94 MW 11,8 SD 1,0	oben	12	12	Σ 90 MW 11,3 SD 1,2	oben			Σ -- MW -- SD --	oben	12	12	Σ 94 MW 11,8 SD 0,5
II		12	12			13	10							12	12	
III		12	11			12	10							12	11	
IV		11	10			11	10							12	11	
V	Mitte	11	9	Σ 74 MW 9,3 SD 1,5	Mitte	11	9	Σ 74 MW 9,3 SD 1,5	Mitte	13	5	Σ 76,5 MW 9,6 SD 2,5	Mitte	11	9	Σ 75,5 MW 9,4 SD 1,2
VI		11	8			11	7			11	8,5			11	8,5	
VII		10	8			10	8			11	7,5			10	8	
VIII		10	7			10	8			11	10			10	8	
IX	unten	9	9	Σ 61 MW 7,6 SD 1,8	unten	9	8	Σ 61,5 MW 7,7 SD 1,2	unten	9,5	9	Σ 18,5 MW 9,3 SD 0,4	unten	9	9	Σ 62 MW 7,8 SD 1,6
X		9	9			8,5	8							9	8	
XI		8	5			8	7							9	7	
XII		7	5			8	5							6	5	
I-XII	alle	N 24	MW 9,5	MW 9,4 SD 1,9	alle	N 24	MW 9,4	MW 9,4 SD 1,9	alle	N 10	MW 9,5	MW 9,5 SD 2,2	alle	N 24	MW 9,6	MW 9,6 SD 2,0
		Σ 229	SD 2,2			Σ 226	SD 1,9			Σ 95	SD 2,2			Σ 232	SD 2,0	
<div><div>Legende: P-Paar Probandenpaar: Proband 1 (P1), Proband 2 (P2)</div><div>Fortsetzung Legende: N Anzahl der Probanden</div></div>																

Abb. A 4: Verteilung der Ergebnisse des NAW-Tests zum Messzeitpunkt 1 der Probandenpaare der Experimentalgruppen

A.3.2

**Vergleich der Ergebnisse des NAW-Tests zum Messzeitpunkt 1 mit den Niveaustufen
des Experimentierens der Probandenpaare der Experimentalgruppen**

P-Paar Nr.:	Kontrollgruppe					TG 1 (Reproduktion)					TG 2 (Reorganisation)					TG 3 (Transfer)				
	MZP 1			NS		MZP 1			NS		MZP 1			NS		MZP 1			NS	
	Bereich / Punkte			Exp.		Bereich / Punkte			Exp.		Bereich / Punkte			Exp.		Bereich / Punkte			Exp.	
	P1		P2	EP1	EP2	P1		P2	EP1	EP2	P1		P2	EP1	EP2	P1		P2	EP1	EP2
I	<i>oben</i>	13	13	2	1	<i>oben</i>	12	12	3	2	<i>oben</i>					<i>oben</i>	12	12	3	2
II		12	12	2	1		13	10	3	2							12	12	3	1
III		12	11	3	1		12	10	3	2							12	11	3	2
IV		11	10	3	2		11	10	3	2							12	11	3	2
V	<i>Mitte</i>	11	9	2	1	<i>Mitte</i>	11	9	3	2	<i>Mitte</i>	13	5	3	2	<i>Mitte</i>	11	9	3	2
VI		11	8	1	1		11	7	3	2		11	8,5	3	2		11	8,5	3	2
VII		10	8	3	1		10	8	3	3		11	7,5	3	1		10	8	3	1
VIII		10	7	2	1		10	8	3	2		11	10	3	2		10	8	2	1
IX	<i>unten</i>	9	9	3	2	<i>unten</i>	9	8	3	2	<i>unten</i>	9,5	9	2	1	<i>unten</i>	9	9	3	1
X		9	9	2	1		8,5	8	1	1							9	8	2	1
XI		8	5	1	1		8	7	2	1							9	7	3	1
XII		7	5	1	1		8	5	3	2							6	5	1	1

Legende:

P-Paar Probandenpaar: Proband 1 (P1), Proband 2 (P2)

EP Experimentierphase: Profilkonstruktion vorbereiten (EP1), Profil konstruieren (EP2)

Niveaustufe (NS) systematisch (Niveaustufe 3), semi-systematisch (Niveaustufe 2), unsystematisch (Niveaustufe 1)

Abb. A 5a: Vergleich der NAW-Tests zum Messzeitpunkt 1 mit den Niveaustufen des Experimentierens der Probandenpaare der Experimentalgruppen

Fortsetzung A.3.2

**Vergleich der Ergebnisse des NAW-Tests zum Messzeitpunkt 2 mit den Niveaustufen
des Experimentierens der Probandenpaare der Experimentalgruppen**

	Kontrollgruppe					TG 1 (Reproduktion)					TG 2 (Reoganisation)					TG 3 (Transfer)				
P-Paar	MZP 2			NS		MZP 2			NS		MZP 2			NS		MZP 2			NS	
Nr.:	Bereich / Punkte			Exp.		Bereich / Punkte			Exp.		Bereich / Punkte			Exp.		Bereich / Punkte			Exp.	
	P1		P2	EP1	EP2	P1		P2	EP1	EP2	P1		P2	EP1	EP2	P1		P2	EP1	EP2
I	oben	13	13	2	1	oben	11	12	3	2	oben					oben	12	13	3	2
II		10	11	2	1		13	11	3	2							12	12	3	1
III		10	12	3	1		13	10	3	2							13	12	3	2
IV		11	12	3	2		11	10	3	2							13	10	3	2
V	Mitte	12	10	2	1	Mitte	11	11	3	2	Mitte	12	4	3	2	Mitte	11	10	3	2
VI		10	8	1	1		11	9	3	2		10	8,5	3	2		11	9	3	2
VII		10	10	3	1		10	9	3	3		12	10,5	3	1		10	10	3	1
VIII		10	7	2	1		11	8	3	2		12	12	3	2		10	10	2	1
IX	unten	11	11	3	2	unten	7	9	3	2	unten	8	8	2	1	unten	11	8,5	3	1
X		7	6,5	2	1		8,5	8	1	1							9	9	2	1
XI		5,5	6	1	1		10	9	2	1							10	11	3	1
XII		7	5	1	1		10	8,5	3	2							8	8	1	1
Legende:																				
P-Paar		Probandenpaar: Proband 1 (P1), Proband 2 (P2)																		
EP		Experimentierphase: Profilkonstruktion vorbereiten (EP1), Profil konstruieren (EP2)																		
Niveaustufe (NS)		systematisch (Niveaustufe 3), semi-systematisch (Niveaustufe 2), unsystematisch (Niveaustufe 1)																		

Abb. A 5b: Vergleich der NAW-Tests zum Messzeitpunkt 2 mit den Niveaustufen des Experimentierens der Probandenpaare der Experimentalgruppen

Fortsetzung A.3.2

**Vergleich der Ergebnisse des NAW-Tests zum Messzeitpunkt 3 mit den Niveaustufen
des Experimentierens der Probandenpaare der Experimentalgruppen**

	Kontrollgruppe				TG 1 (Reproduktion)				TG 2 (Reoganisation)				TG 3 (Transfer)							
P-Paar	MZP 3			NS		MZP 3			NS		MZP 3			NS		MZP 3			NS	
Nr.:	Bereich / Punkte			Exp.		Bereich / Punkte			Exp.		Bereich / Punkte			Exp.		Bereich / Punkte			Exp.	
	P1		P2	EP1	EP2	P1		P2	EP1	EP2	P1		P2	EP1	EP2	P1		P2	EP1	EP2
I	oben	9,5	13	2	1	oben	13	12	3	2	oben					oben	10	12	3	2
II		12	11	2	1		10	11	3	2							12	10	3	1
III		10	11	3	1		12	11	3	2							12	13	3	2
IV		11	11	3	2		11	11	3	2							12,5	10	3	2
V	Mitte	11	12	2	1	Mitte	11	10	3	2	Mitte	13	6	3	2	Mitte	10	8	3	2
VI		11	11	1	1		10	9	3	2		10	9,5	3	2		11	10	3	2
VII		10	9,5	3	1		12	11	3	3		13	9,5	3	1		10	10	3	1
VIII		11	5,5	2	1		12	10	3	2		13	12	3	2		11	11	2	1
IX	unten	13	13	3	2	unten	8,5	11	3	2	unten	9	9	2	1	unten	13	8	3	1
X		10	7	2	1		11,5	9	1	1							6	10	2	1
XI		3,5	7,5	1	1		9	9	2	1							10	10	3	1
XII		7	7	1	1		10	9,5	3	2							3	4	1	1
Legende:																				
P-Paar		Probandenpaar: Proband 1 (P1), Proband 2 (P2)																		
EP		Experimentierphase: Profilkonstruktion vorbereiten (EP1), Profil konstruieren (EP2)																		
Niveaustufe (NS)		systematisch (Niveaustufe 3), semi-systematisch (Niveaustufe 2), unsystematisch (Niveaustufe 1)																		

Abb. A 5c: Vergleich der NAW-Tests zum Messzeitpunkt 3 mit den Niveaustufen des Experimentierens der Probandenpaare der Experimentalgruppen

Abbildungsverzeichnis

Abb. 1	Die Dimensionierung und Graduierung des offenen Experimentierens nach dem eigenen Ansatz im Überblick.....	18
Abb. 2	Design der explorativen Vorstudie I.....	44
Abb. 3	Design der explorativen Vorstudie II.....	54
Abb. 4	Design der Hauptstudie.....	68
Abb. 5	Strukturierung des Messinstruments Beobachtungsbogen	83
Abb. 6	Prozessdiagramme des Referenzbeobachters (Video) zur Ermittlung der Beobachterübereinstimmung.....	98
Abb. 7	Gruppenvergleiche für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests zum Messzeitpunkt 1 (Vortest). Wertebereich 0 – 13 Punkte.....	105
Abb. 8a	Prozessgrafik eines Probandenpaares der Testgruppe 1 (Reproduktion).....	115
Abb. 8b	Fortsetzung der Prozessgrafik eines Probandenpaares der Testgruppe 1 (Reproduktion).....	115
Abb. 9a	Prozessgrafik eines Probandenpaares der Kontrollgruppe	118
Abb. 9b	Fortsetzung der Prozessgrafik eines Probandenpaares der Kontrollgruppe.....	118
Abb. 10	Niveaustufengrafik von Repräsentantenpaaren der Experimentiertypen A – D.....	125
Abb. 11	Verteilung der Experimentiertypen A – D.....	126
Abb. 12	Verteilung der Niveaustufen des Experimentierens in Abhängigkeit der Experimentierphasen und Kontexte der vorbereitenden Instruktion.....	134

Abbildungen des Anhangs

Abb. A 1	Experimentieranordnung zur Anströmung von Windradprofilen zur Aufnahme von Messfolgen für die vorbereitende Instruktion der Testgruppen TG (Vorstudie) und TG 1 (Hauptstudie) mit der Transferleistung Reproduktion sowie zur Bearbeitung der offenen Experimentieraufgabe.	178
Abb. A 2	Experimentieranordnung zur Anströmung von Tragflächenprofilen zur Aufnahme von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 2 (Hauptstudie) mit der Transferleistung Reorganisation.	178
Abb. A 3	Experimentieranordnung zum Abwurf von Wurfkörpern zur Aufnahme von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 3 (Hauptstudie) mit der Transferleistung Transfer.	178
Abb. A 4	Verteilung der Ergebnisse des NAW-Tests zum Messzeitpunkt 1 der Probandenpaare der Experimentalgruppen	225
Abb. A 5a	Vergleich der Ergebnisse des NAW-Tests zum Messzeitpunkt 1 mit den Niveaustufen des Experimentierens der Probandenpaare der Experimentalgruppen	226
Abb. A 5b	Vergleich der Ergebnisse des NAW-Tests zum Messzeitpunkt 2 mit den Niveaustufen des Experimentierens der Probandenpaare der Experimentalgruppen.	227
Abb. A 5c	Vergleich der Ergebnisse des NAW-Tests zum Messzeitpunkt 3 mit den Niveaustufen des Experimentierens der Probandenpaare der Experimentalgruppen.	228

Tabellenverzeichnis

Tab. 1	Ergebnisse der ersten empirisch-quantitativen Exploration: Taxonomie der Planungsschritte.	48
Tab. 2	Fähigkeitsbereiche und Teilfähigkeiten des Konstrukts „Allgemeine Fähigkeiten im Umgang mit Variablen“	76
Tab. 3a	Numerisches Codiersystem zur Codierung beobachtbarer Experimentiertätigkeiten in Experimentierphase 1.	89
Tab. 3b	Numerisches Codiersystem zur Codierung beobachtbarer Experimentiertätigkeiten in Experimentierphase 2.	90
Tab. 4a	Kategorien der verbalen Argumentation „Vergleich und Bewertung von Messfolgen“	93
Tab. 4b	Kategorien der verbalen Argumentation „Strategie und Idee zur Profilkonstruktion“	93
Tab. 4c	Kategorien der verbalen Argumentation „Erklärung“	94
Tab. 5	Anzahl der erfassten verbalen Argumentationen zur Ermittlung der Beobachterübereinstimmung.	99
Tab. 6	Kolmogorow-Smirnow-Anpassungstest (KS-Test) auf Normalverteilung für abhängige Stichproben mit Differenzen der Messwertpaare im NAW-Test zu den Messzeitpunkten. Fettdruck: signifikante Abweichung mit $d > d_{krit}$	103
Tab. 7	t-Tests auf Unterschiede zwischen den Messzeitpunkten 1 und 2 sowie 2 und 3 der Kontroll- und Testgruppen für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests. Fettdruck: signifikante Effekte für $p < 0,05$	106
Tab. 8a	Korrelationskoeffizienten r nach Pearson zwischen den Messzeitpunkten 1 – 3 der Experimentalgruppen für die	

	Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests.	111
Tab. 8b	Signifikante Korrelationen des Korrelationskoeffizienten r nach Pearson zwischen den Messzeitpunkten 1 – 3 der Experimental- gruppen für die Dimension Naturwissenschaftliche Arbeitsweisen (NAW) des Fach- & NAW-Tests. Fettdruck: signifikante Effekte für $p < 0,05$	111
Tab. 9	Kriterien und Kategorien zur Bewertung und Einweisung der Experimentierprozesse in Niveaustufen	123
Tab. 10	Charakteristik der Experimentiertypen A – D.	125
Tab. 11	Verteilung der Experimentiertypen A – D.	126
Tab. 12	Verteilung der Effekte des Experimentiertyps C.	131
Tab. 13	Verteilung der Effekte des Experimentiertyps D.	133
Tab. 14	Verteilung der Niveaustufenkombinationen	138
Tab. 15	Mann-Whitney U-Tests mit Rangbindung zwischen Test- und Kontrollgruppen für Experimentierphase 1. Fettdruck: signifikante Effekte für $p < 0,05$	141
Tab. 16	Mann-Whitney U-Tests mit Rangbindung zwischen Test- und Kontrollgruppen für Experimentierphase 2. Fettdruck: signifikante Effekte für $p < 0,05$	142
Tab. 17	Mann-Whitney U-Tests mit Rangbindung zwischen Test- und Kontrollgruppen mit Aufhebung der Experimentier- phasentrennung zwischen Experimentierphase 1 und 2. Fettdruck: signifikante Effekte für $p < 0,05$	143
Tab. 18	Wilcoxon Vorzeichen-Rangtests – mit gebundenen (betragsgleichen) Differenzen – auf Niveaustufenunter- schiede des Experimentierens zwischen den Experimentier- phasen 1 und 2 für die Test- und Kontrollgruppen. Fettdruck: signifikante Effekte für $p < 0,05$	144
Tab. 19	Rangkorrelationskoeffizient r_s nach Spearman zwischen den Ergebnissen der NAW-Tests zu den Messzeitpunkten 1 – 3 und den Niveaustufen des Experimentierens für die Experimentierphasen 1 und 2 der Teilstichprobe. Fettdruck: signifikante Korrelationen für $p < 0,05$	149

Tabellen des Anhangs

Tab. A 1	Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppen TG (Vorstudie) und TG 1 (Hauptstudie) mit der Transferleistung Reproduktion.	173
Tab. A 2a	Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 2 (Hauptstudie) mit der Transferleistung Reorganisation	174
Tab. A 2b	Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 2 (Hauptstudie) mit der Transferleistung Reorganisation	175
Tab. A 3a	Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 3 (Hauptstudie) mit der Transferleistung Transfer.	176
Tab. A 3b	Messergebnisse der Voruntersuchung zur Auswahl von Messfolgen für die vorbereitende Instruktion der Testgruppe TG 3 (Hauptstudie) mit der Transferleistung Transfer.	177
Tab. A 4	Messinstrument Fach- & NAW-Test: Aufbau – Struktur – Bewertung.....	203

Literaturverzeichnis

- Baumert, J., Lehmann, R., Lehrke, M., Clausen, M., Hosenfeld, I., Neubrand, J., Patjens, S., Jungclaus, H., Günther, W. (1998). *Testaufgaben Naturwissenschaften TIMSS 7./8. Klasse (Population 2). Materialien aus der Bildungsforschung, Nr. 61.* Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Bos, W., Klieme, E., Lehmann, R., Lehrke, M., Hosenfeld, I., Neubrand, J., Watermann, R. (1999). *Testaufgaben zu TIMSS/III. Mathematisch-naturwissenschaftliche Grundbildung und voruniversitäre Mathematik und Physik der Abschlußklassen der Sekundarstufe II (Population 3). Materialien aus der Bildungsforschung, Nr. 62.* Berlin: Max-Planck-Institut für Bildungsforschung.
- Baxter, G. P. & Shavelson, R. J. (1994). Science Performance Assessments: Benchmarks and Surrogates. *International Journal of Educational Research*, 21/3, 279–298.
- Berg, C., Bergendahl, V. & Lundberg, B. (2003). Benefiting from an open-ended experiment? A comparison of attitudes to, and outcomes of, an expository versus an open-inquiry version of the same experiment. *International Journal of Science Education*, 25/3, 351–372.
- Braun, T. (2009). *Offene Experimente in der Lehramtsausbildung. Analyse physikalischer Phänomene für eine naturwissenschaftliche Experimentierweise.* Dissertation an der Universität Duisburg-Essen. Eingesehen unter <http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=20589> [25.10.2011]
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Berlin: Springer.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler.* Heidelberg: Springer.
- Buck, L. B., Bretz, S. L. & Towns, M. H. (2008). Characterizing the Level of Inquiry in the Undergraduate Laboratory. *Journal of College Science Teaching*, September/October, 52–58.

- Bullock, M. & Ziegler, A. (1993). Scientific Thinking. In F. E. Weinert & W. Schneider (Hrsg.). *LOGIK Report No. 10: Assessment, procedures and results of Wave 7* (pp. 59–89). Munich: Max-Planck-Institute for Psychological Research.
- Bullock, M. & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (Eds.). *Individual development from 3 to 12. Findings from the Munich Longitudinal Study* (pp. 38–54). Cambridge: University Press.
- Bullock, M. & Sodian, B. (2003). Entwicklung des wissenschaftlichen Denkens. In W. Schneider und M. Knopf (Hrsg.). *Entwicklung, Lehren und Lernen. Zum Gedenken an Franz Emanuel Weinert* (S. 75–92). Göttingen: Hogrefe.
- Burns, J. C., Okey, J. R., Wise, K. C. (1985). Development of an Integrated Process Skill Test: TIPS II. *Journal of Research in Science Teaching*, 22/2, 169–177.
- Bybee, R. (1997). *Achieving scientific literacy: from purposes to practices*. Portsmouth: Heinemann.
- Carey, S., Evans, R., Honda, M., Jay, E. & Unger, C. (1989). An experiment is when you try it and see if it works'. A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11/5, 514–529.
- Chen, Z. & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70/5, 1098–1120.
- Chinn, C. A. & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, 35/6, 623–654.
- Diehl, J., Arbinder, R. (2001). *Einführung in die Inferenzstatistik*. Eschborn bei Frankfurt a. M.: Dietmar Klotz.
- Diekmann, A. (2007). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt.
- Domin, D. S. (1999). A Review of Laboratory Instruction Styles. *Journal of Chemical Education*, 76, 543–547.
- Domin, D. S. (2007). Students' perceptions of when conceptual development occurs during laboratory instruction. *Chemistry Education Research and Practice*, 8/2, 140–152.
- Dubs, F. (1966). *Aerodynamik der reinen Unterschallströmung*. Basel: Birkhäuser.
- Duit, R. (1999). TIMSS-Items für die Klassenstufen 7 und 8. *Naturwissenschaften im Unterricht Physik*, 10/54, 236–242.

- Duit, R. (2003). Naturwissenschaftliches Arbeiten. *Naturwissenschaften im Unterricht Physik*, 74/14, 4–8.
- Duit, R., Gropengießer, H. & Stäudel, L. (2007). Naturwissenschaftliches Arbeiten. Eine Einführung. In R. Duit, H. Gropengießer & L. Stäudel (Hrsg.). *Naturwissenschaftliches Arbeiten. Unterricht und Material 5–10* (S. 4–8). Seelze: Friedrich.
- Edgeworth, M. & Edgeworth, R. L. (1811). *Essays on Practical Education*. London: Johnson.
- Emden, M. (2011). *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens. Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*. Berlin: Logos.
- Emden, M. & Sumfleth, E. (2012). Prozessorientierte Leistungsmessung. Zur Eignung einer Protokollmethode für die Bewertung von Experimentierprozessen. *Der mathematische und naturwissenschaftliche Unterricht*, 65/2, 68–75.
- Engeln, K. (2004). *Schülerlabors: authentische, aktivierende Lernumgebungen als Möglichkeit, Interesse an Naturwissenschaften und Technik zu wecken*. Berlin: Logos.
- Fischer, H. E. & Draxler, D. (2001). Aufgaben und naturwissenschaftlicher Unterricht. *Der mathematische und naturwissenschaftliche Unterricht*, 54/7, 388–393.
- Forschergruppe Kassel (2006). Archimedes und die Sache mit der Badewanne. In G. Becker, M. Horstkemper, E. Risse, L. Stäudel, R. Wernig & F. Winter (Hrsg.). *Diagnostizieren und Fördern. Stärken entdecken – Können entwickeln* (S. 84–88). Friedrich Jahresheft XXIV. Seelze: Erhard Friedrich.
- Forschergruppe Kassel (2007). Schritt für Schritt zur Lösung. Differenzierung durch Aufgaben mit gestuften Lernhilfen. *Naturwissenschaften im Unterricht Physik*, 18/99–100, 42–45.
- Friedrichs, J. (1990). *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag.
- Fröhlich, I., Bieber, G. & Horn, M. E. (2004). Physiklernen mit offenen Aufgaben – Erfahrungen mit dem BLK-Programm SINUS Brandenburg. *Frühjahrstagung der Deutschen Physikalischen Gesellschaft 2003 in Augsburg*. CD-ROM. Berlin: Lehmann's.

- Glomski, J. (2008). *Ausarbeitung von Experimenten zur Umströmung von Tragflächenmodellen für Projekte im Schülerlabor*. Masterarbeit an der Ruhr-Universität Bochum.
- Gräber, W., Nentwig, P., Koballa, T. & Evans, R. (2002). *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. Opladen: Leske und Budrich.
- Hagemeister, V. (1999). Was wurde bei TIMSS erhoben?. *Die Deutsche Schule*, 91, 160–177.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. *Der mathematische und naturwissenschaftliche Unterricht*, 57/4, 196–203.
- Hammann, M., Phan, T. T. H., Ehmer, M. & Bayrhuber, H. (2006). Fehlerfrei Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht*, 59/5, 292–299.
- Hammann, M. (2006a). PISA und Scientific Literacy. In U. Steffens und R. Messner (Hrsg.). *PISA macht Schule. Konzeptionen und Praxisbeispiele zur neuen Aufgabenkultur* (S. 127–179). Wiesbaden: Institut für Qualitätsentwicklung.
- Hammann, M. (2006b). Kompetenzförderung und Aufgabenentwicklung. *Der mathematisch und naturwissenschaftliche Unterricht*, 59/2, 85–95.
- Hammann, M. (2007). Das Scientific Discovery as Dual Search-Modell. In D. Krüger & H. Vogt (Hrsg.). *Theorien in der biologiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 187–196). Berlin: Springer.
- Hammann, M., Ganser, M. & Haupt, M. (2007). Experimentieren können: Kompetenzentwicklungsmodelle und ihre Nutzung im Unterricht. *Geographie heute*, 255/ 266, 88–91.
- Hammann, M., Phan, T. T. H., Bayrhuber, H. (2007). Experimentieren als Problemlösen: Lässt sich das SDDS-Modell nutzen, um unterschiedliche Dimensionen beim Experimentieren zu messen?. *Zeitschrift für Erziehungswissenschaft*, 10, Sonderheft 8, 33–49.
- Hammann, M., Phan, T. T. H., Ehmer, M., Bayrhuber, H. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education*, 42/2, 66–72.
- Harmon, M., Smith, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V. S. et al. (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

- Heege, R. (1978). Begriffsbestimmung der Didaktik der Physik und Konsequenzen für die Offenheit des Lernens: Über Improvisation I, II. *physica didact.* 5, Sonderheft Nr. 1, 49–58, 59–64.
- Heege, R. & Bruns, B. (1981). Improvisation und Experiment. *Physik und Didaktik* 9/1, 20–29.
- Hegarty, E. (1978). Levels of scientific enquiry in university science laboratory classes: implications for curriculum deliberations. *Research in Science Education*, 8, 45–57.
- Heinzel, H. (2006). *Empirische Fallstudie und vergleichender Sachstandsbericht zur Offenheit in Unterricht, Schule und deren Umfeld*. Dissertation an der Universität Siegen. Eingesehen unter <http://dokumentix.ub.uni-siegen.de/opus/volltexte/2007/226> [25.10.2011]
- Hemlo-Silver, C. E., Duncan, R. G. & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 43, 99–107.
- Herron, M. D. (1971). The Nature of Scientific Inquiry. *The School Review*, 79/2, 171–212.
- Hodson, D. (1992). Assessment of Practical Work: Some Considerations in Philosophy of Science. *Science and Education*, 1/2, 115–144.
- Hodson, D. (1993). Re-thinking old ways: Towards a more critical approach to practical work in school science. *Studies in Science Education*, 22, 85–142.
- Hof, S. (2010). *Wissenschaftsmethodischer Kompetenzerwerb durch Forschendes Lernen. Entwicklung und Evaluation einer Interventionsstudie*. Kassel: University Press.
- Hopf, M. (2007). *Problemorientierte Schülerexperimente*. Berlin: Logos.
- Johnstone, S. H. & Wham, A. J. B. (1982). The demands of practical work. *Education in Chemistry*, 19/3, 71–73.
- Jong, T. de & Joolingen, W. R. van (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68/2, 179–201.
- Joolingen, W. R. van & Jong, T. de (1997). An extended dual search space model of scientific discovery learning. *Instructional Science*, 25/5, 307–346.
- Kanari, Z. & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, 41/7, 748–769.

- Kirchner, S. & Priemer, B. (2007). Probleme von Schülern mit Offenen Experimentieraufgaben in Physik. In D. Höttecke (Hrsg.). *Naturwissenschaftlicher Unterricht im internationalen Vergleich. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung der GDGP in Bern 2006* (S. 346–348). Berlin: LIT.
- Kirchner, S. & Priemer, B. (2008). Wie können Schüler angeleitet werden, Experimente selber zu planen? In D. Höttecke (Hrsg.). *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Essen 2007* (S. 194–196). Berlin: LIT.
- Kirchner, S. & Priemer, B. (2010). Welche Kompetenzen zeigen Schüler beim Umgang mit Variablen? In D. Höttecke (Hrsg.). *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Dresden 2009* (S. 206–208). Berlin: LIT.
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-based, Experimental, and Inquiry-based Teaching. *Educational Psychologist*, 41, 75–86.
- Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12/1, 1–48.
- Klahr, D. (2000). *Exploring science. The cognition and development of discovery processes*. Cambridge, Massachusetts: MIT.
- Klahr, D. & Nigam, M. (2004). The equivalence of learning paths in early science instruction. Effects of direct instruction and discovery learning. *Psychological Science*, 15/10, 661–667.
- Klayman, J. & Ha, Y.-W. (1989). Hypothesis Testing in Rule Discovery: Strategy, Structure, and Content. *Journal of Experimental Psychology*, 15/4, 596–604.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruper, H., Prenzel, P. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards – Expertise*. Berlin: BMBF.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. In M. Prenzel und L. Allolio-Näcke (Hrsg.). *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG – Schwerpunktprogramms* (S. 127–146). Münster: Waxmann.
- Klopfer, E. (1971). Evaluation of learning in science. In B. Bloom (Ed.). *Handbook on formative and summative evaluation of student learning* (pp. 599–641). New York: Mc Graw Hill.

- Klopfer, L. E. (1990). Learning Scientific Enquiry in the Student Laboratory. In Hegarty-Hazel, E. (Ed.). *The Student Laboratory and the Science Curriculum*. London: Routledge.
- Klos, S., Henke, C., Kieren, C., Walpuski, M. & Sumfleth, E. (2009). Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei unterschiedliche Kompetenzen. *Zeitschrift für Pädagogik*, 3, 304–321.
- Korneck, F. (1998). *Die Strömungsdynamik als Zugang zur nichtlinearen Dynamik. Entwicklung, Erprobung und Evaluation einer Unterrichtsreihe für die gymnasiale Oberstufe und die Lehrerbildung*. Aachen: Shaker.
- Kühnelt, H. (1999). TIMSS-Items für die Sekundarstufe II. *Naturwissenschaften im Unterricht Physik*, 10/ 54, 243–248.
- Kühnelt, H. (2000). TIMSS in der Mittelstufe. *Plus Lucis. Mitteilungsblatt des Vereins zur Förderung des Physikalischen und Chemischen Unterrichts und des Fachausschuß LHS der Österreichischen Physikalischen Gesellschaft*, 1, 9–12.
- Küsting, J., Thilmann, H., Wirth, J., Fischer, H.-E. & Leutner, D. (2008). Strategisches Experimentieren im naturwissenschaftlichen Unterricht. *Psychologie in Erziehung und Unterricht*, 55, 1–15.
- Kuhn, D. & Angelev, J. (1976). An experimental study of the development of formal operational thought. *Child Development*, 47/3, 697–706.
- Labudde, P. & Stebler, R. (1999). Lern- und Prüfungsaufgaben für den Physikunterricht. Erträge aus dem TIMSS-Experimentiertest. *Naturwissenschaften im Unterricht Physik*, 10/54, 23–31.
- Labudde, P., Metzger, S. & Gut, C. (2009). Bildungsstandards: Validierung des Schweizer Kompetenzmodells. In D. Höttercke (Hrsg.). *Chemie- und Physikdidaktik für die Lehramtsausbildung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Schwäbisch Gmünd 2008* (S. 307–308). Berlin: LIT.
- Lawson, A. E. & Wollmann, W. T. (1976). Encouraging the transition from concrete to formal cognitive function – an experiment. *Journal of Research in Science Teaching*, 13/5, 413–430.
- Leach, J. & Paulsen, A. (1999). *Practical Work in Science Education*. Dordrecht: Kluwer.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.). *PISA 2003. Der Bildungsstandard der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 147–175). Münster: Waxmann.

- Lewis, J. (2002). The Effectiveness of Mini-Projects as a Preparation for open-ended Investigations. In D. Psillos & H. Niedderer (Eds.). *Teaching and Learning in the Science Laboratory* (pp. 139–150). Dordrecht: Kluwer.
- Lienert, G. A., Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lin, X. & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, 36/7, 837–858.
- Lunetta, V. N., Hofstein, A. & Clough, M. P. (2007). Learning and Teaching in the School Science Laboratory: An Analysis of Research, Theory, and Practice. In S. K. Abell & N. G. Lederman (Eds.). *Handbook of Research on Science Education* (pp. 393–441). London: Lawrence Erlbaum.
- Mannel, S., Sumfleth, E. & Walpuski, M. (2009). Schwierigkeitsbestimmende Faktoren von Aufgaben zu experimentell-naturwissenschaftlichen Arbeitsweisen für den unteren Leistungsbereich. In D. Höttecke (Hrsg.). *Chemie- und Physikdidaktik für die Lehramtamsausbildung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Schwäbisch Gmünd 2008* (S. 380–382). Berlin: LIT.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59/1, 14–19.
- Mayer, J. & Ziemek, H.-P. (2006). Offenes Experimentieren. Forschendes Lernen im Biologieunterricht. *Unterricht Biologie*, 317, 4–12.
- Mayer, J., Grube, C. & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms & A. Sandmann (Hrsg.). *Lehr- und Lernforschung in der Biologiedidaktik, Band 3. Ausbildung und Professionalisierung von Lehrkräften. Internationale Tagung der Fachsektion Didaktik der Biologie im VBiO, Essen 2007* (S. 63–80). Innsbruck: Studienverlag.
- Merzyn, G. (1987). Die Sprache unserer Schulbücher. *Der mathematische und naturwissenschaftliche Unterricht*, 40/2, 75–80.
- Merzyn, G. (1994). *Physikschulbücher, Physiklehrer und Physikunterricht*. Kiel: IPN.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.) (2008). *Kernlehrplan für das Gymnasium – Sekundarstufe I in Nordrhein-Westfalen. Physik*. Frechen: Ritterbach.

- Mikelskis-Seifert, S. & Rabe, T. (2007). *Physik Methodik. Handbuch für die Sekundarstufe I und II*. Berlin: Cornelsen.
- Millar, R. (1998). Rhetoric and Reality: What practical work in science education is really for. In Wellington, J. (Ed.). *Practical Work in School Science: Which Way now?*. London: Routledge.
- Millar, R., Tiberghien, A. & Le Maréchal, J.-F. (2002). Varieties of Labwork: A Way of Profiling Labwork Tasks. In D. Psillos & H. Niedderer (Eds.). *Teaching and Learning in the Science Laboratory*. Dordrecht: Kluwer.
- Moosbrugger, H., Kelava, A. (2008). *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer.
- Mothes, H. (1968). *Methodik und Didaktik der Physik und Chemie*. Köln: Aulis.
- Nawrath, D., Maisyenka, V. & Schecker, H. (2011). Experimentelle Kompetenz. Ein Modell für die Unterrichtspraxis. *Praxis der Naturwissenschaften – Physik in der Schule*, 60/6, 42–49.
- Okada, T. & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science*, 21/2, 109–146.
- Pauli, C. & Reusser, K. (2003). Unterrichtsskripts im schweizerischen und im deutschen Mathematikunterricht. *Unterrichtswissenschaften*, 31, 238–272.
- Peschel, F. (2002). Offener Unterricht: Idee – Realität – Perspektive und ein praxiserprobtes Konzept zur Diskussion. Hohengehren: Schneider.
- Pinther, A. (1980). Beobachtung. In Friedrich, W. & Henning, W. (Hrsg.). *Der sozialwissenschaftliche Forschungsprozeß*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- PISA–Beispielaufgaben aus dem Naturwissenschaftstest (2006). Abgerufen unter http://pisa.ipn.uni-kiel.de/PISA06_Science_Beispielaufgaben.pdf [08.01.2008]
- PISA–Lösungen der Beispielaufgaben aus dem Naturwissenschaftstest (2006). Abgerufen unter http://pisa.ipn.uni-kiel.de/PISA06_Science_Beispielaufgaben_Loesungen.pdf [08.01.2008]
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., Pekrun, R. (2007). *Pisa 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Priemer, B. & Kirchner, S. (2007a). Offenes Experimentieren mit Windrädern – Konstruktion von Rotorblättern zur Spannungs- und Leistungsoptimierung. *Praxis der Naturwissenschaften – Physik in der Schule*, 56/3, 11–16.

- Priemer, B. & Kirchner, S. (2007b). Students' attitudes toward open inquiry experiments in physics. In P. J. Gilmer, J. Osborne, J. Shymansky & J. Tillotson (Eds.). *National Association for Research in Science Teaching, Annual Conference 2007, New Orleans, Restructuring Science Education Through Research*, CD-ROM.
- Priemer, B. & Kirchner, S. (2007c). Einstellungen von Schülern zu offenen Experimentieraufgaben in Physik. In D. Höttecke (Hrsg.). *Naturwissenschaftlicher Unterricht im internationalen Vergleich. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung der GDChP in Bern 2006* (S. 343–346). Berlin: LIT.
- Priemer, B. & Kirchner, S. (2008). Welche Kenntnisse haben Schüler, um Experimente selber zu entwerfen? In D. Höttecke (Hrsg.). *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Essen 2007* (S. 296–298). Berlin: LIT.
- Priemer, B., Uhlmann, S. & Kirchner, S. (2010). Was ist eigentlich offen beim offenen Experimentieren? In D. Höttecke (Hrsg.). *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Dresden 2009* (S. 239–241). Berlin: LIT.
- Priemer, B. (2011). Was ist das Offene beim offenen Experimentieren?. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 315–337.
- Psillos, D. & Niedderer, H. (2002). *Teaching and learning in the science laboratory*. Dordrecht: Kluwer.
- Rasch, B., Frieze, M. (2006). *Quantitative Methoden. Einführung in die Statistik*. Heidelberg: Springer.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2006). *Quantitative Methoden 2*. Heidelberg: Springer.
- Reinhold, P. (1996). *Offenes Experimentieren und Physiklernen*. Kiel: IPN.
- Reinhold, P. (1997a). Offenes Experimentieren als Lernform. In H. Behrenndt (Hrsg.). *Zur Didaktik der Physik und Chemie. Vorträge auf der Tagung für Didaktik der Physik/Chemie in Bremen 1996* (S. 41–55). Alsbach: Leuchtturm.
- Reinhold, P. (1997b). Offenes Experimentieren: Ein neuer Ansatz für den Physikunterricht?. In H. E. Fischer (Hrsg.). *Handlungsorientierter Physik – Unterricht Sekundarstufe II*. Bonn: Dümmler.
- Reyer, T., Trendel, G. & Fischer, H. E. (2004). Was kommt beim Schüler an? – Lehrerintentionen und Schülerlernen im Physikunterricht. In J. Doll und M.

- Prenzel (Hrsg.). *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 195–211). Münster: Waxmann.
- Rost, D. H. (2007). *Interpretation und Bewertung pädagogisch-psychologischer Studien*. Weinheim: Beltz.
- Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching*, 33/10, 1045–1063.
- Schauble, L., Klopfer, L. E. & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28/9, 859–882.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32/1, 102–119.
- Schecker, H., Fischer, H. E. & Wiesner, H. (2004). Physikunterricht in der gymnasialen Oberstufe. In H.-E. Tenorth (Hrsg.). *Kerncurriculum Oberstufe II* (S. 148–234). Weinheim: Beltz.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Schreiber, N., Theyßen, H. & Schecker, H. (2009). Experimentelle Kompetenz messen?!. *Physik und Didaktik in Schule und Hochschule*, 8/3, 92–101.
- Schreiber, N., Theyßen, H. & Schecker, H. (2011). Diagnostik experimenteller Kompetenz: Ergebnisse von Pilotstudien. In D. Höttecke (Hrsg.). *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Potsdam 2010* (S. 244–246). Berlin: LIT.
- Schröder, H. (1992). *Grundwortschatz Erziehungswissenschaft: ein Wörterbuch der Fachbegriffe von „Abbilddidaktik“ bis „Zielorientierung“*. München: Ehrenwirth.
- Schwab, J. J. (1962). The Teaching of Science as Enquiry. In J. J. Schwab & P. F. Brandwein (Eds.). *The Teaching of Science*. Cambridge, MA: Harvard University Press.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.) (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München: Wolters Kluwer.

- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.) (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München: Wolters Kluwer.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.) (2005c). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München: Wolters Kluwer.
- Séré, M.-G. (2002). Towards Renewed Research Questions from the Outcomes of the European Project. Labwork in Science Education. *Science Education*, 86, 624–644.
- Shavelson, R. J., Ruiz-Primo, M. A. & Wiley, E. W. (1999). Note on sources of sampling variability in science performance assessments. *Journal of Educational Measurement*, 36/1, 61–71.
- Singer, S. R., Hilton, M. L., Schweingruber, H. A. (2006). *America's Lab Report: Investigations in High School Science*. Washington, D. C.: The National Academies Press.
- Sodian, B., Jonen, A., Thoermer, C. & Kircher, E. (2006). Die Natur der Naturwissenschaften verstehen. Implementierung wissenschaftstheoretischen Unterrichts in der Grundschule. In M. Prenzel & L. Allolio-Näcke (Hrsg.). *Untersuchungen zur Bildungsqualität von Schule – Abschlussbericht des DFG – Schwerpunktprogramms* (S. 147–160). Münster: Waxmann.
- Sodian, B., Bullock, M. & Koerber, S. (2008). Wissenschaftliches Denken und Argumentieren. Was muss Hänschen lernen, damit aus Hans etwas wird? In W. Schneider (Hrsg.). *Entwicklung von der Kindheit bis zum Erwachsenenalter. Befunde der Münchener Längsschnittstudie LOGIK* (S. 67–84). Weinheim: Beltz.
- Staer, H., Goodrum, D. & Hackling, M. (1998). High School Laboratory Work in Western Australia: Openness to Inquiry. *Research in Science Education*, 28/2, 219–228.
- Stäudel, L., Franke-Braun, G. & Schmidt-Weigand, F. (2007). Komplexität erhalten – auch in heterogenen Lerngruppen: Aufgaben mit gestuften Lernhilfen. *Chemie konkret*, 14/3, 115–122.
- Stebler, R., Reusser, K. & Ramseier, E. (1998). Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen. Erträge aus dem TIMSS-Experimentiertest. *Bildungsforschung und Bildungspraxis. Schweizerische Zeitschrift für Erziehungswissenschaften*, 20/1, 28–54.

- Thillmann, H., Küsting, J., Wirth, J. & Leutner, D. (2009). Is it merely a question of 'what' to prompt or also 'when' to prompt? – The role of presentation time of prompts in self-regulated learning. *Zeitschrift für Pädagogische Psychologie*, 23, 105–115.
- Tiberghien, A., Veillard, L., Le Maréchal, J.-F., Buty, C. & Millar, R. (2001). An Analysis of Labwork Tasks Used in Science Teaching at Upper Secondary School and University Level in Several European Countries. *Science Education*, 85, 483–508.
- TIMSS Population 2 Item Pool. *TIMSS Science Items: Released Set for Population 2*. Abgerufen unter <http://timss.bc.edu> [14.01.2009]
- TIMSS Population 3 Item Pool. *TIMSS Released Physics Items Population 3*. Abgerufen unter <http://timss.bc.edu> [14.01.2009]
- Trendel, G. & Fischer, H. E. (2007). Naturwissenschaftliche Arbeitsweisen. *Der mathematische und naturwissenschaftliche Unterricht*, 60/7, 388–394.
- Trendel, G., Wackermann, R. & Fischer, H. E. (2007). Lernprozessorientierte Lehrerfortbildung in Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 9–31.
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51/1, 1–10.
- Wahser, I. & Sumfleth, E. (2008). Training experimenteller Arbeitsweisen zur Unterstützung kooperativer Kleingruppenarbeit im Fach Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, 14, 219–241.
- Walpuski, M. (2006). *Optimierung von experimenteller Kleingruppenarbeit durch Strukturierungshilfen und Feedback*. Berlin: Logos.
- Walpuski, M., Kampa, N., Kauertz, A. & Wellnitz, N. (2008). Evaluation der Bildungsstandards in den Naturwissenschaften. *Der mathematische und naturwissenschaftliche Unterricht*, 61/6, 323–326.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, 12/3, 129–140.
- Wieczorek, R. & Sommer, K. (2008). Abgestufte Lernhilfen und deren Einsatz bei eigenverantwortlichen Schülerexperimenten. *Unterricht Chemie*, 106/107, 88–93.
- Wissenschaftliches Konsortium HarmoS Naturwissenschaften+ (2008). *HarmoS Naturwissenschaften+: Kompetenzmodell und Vorschläge für Bildungsstandards*. Bern (Wissenschaftlicher Schlussbericht).

- Weltner, K. (2000). Physikalische Erklärung des aerodynamischen Auftriebs – einfach, überzeugend und kohärent. *Physik in der Schule*, 38/6, 363–368.
- Weltner, K. (2002). Flugphysik – um das Fliegen zu verstehen. *Der mathematische und naturwissenschaftliche Unterricht*, 55/7, 388–396.
- Wodzinski, R. (1999). Wie erklärt man das Fliegen in der Schule? Versuch einer Analyse verschiedener Erklärungsmuster. *Plus Lucis*, 2, 18–22.
- Wodzinski, R. & Ziegler, A. (2000). Erklärung des Fliegens in der Schule. Eine kritische Analyse gängiger Erklärungsmuster und ein alternativer Unterrichtsvorschlag. *Der mathematische und naturwissenschaftliche Unterricht*, 53/5, 273–281.
- Zeyer, A., Adamina, M., Gings, F. & Labudde, P. (2008). HarmoS Naturwissenschaften – Entwicklung, Umsetzung und Assessment von Standards im naturwissenschaftlichen Unterricht der Schweiz. In D. Höttecke (Hrsg.). *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Essen 2007* (S. 206–208). Berlin: LIT.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27/2, 172–223.

Danksagung

Ich danke ...

- Herrn Prof. Dr. Burkhard Priemer für die Möglichkeit, diese Arbeit am Lehrstuhl für Didaktik der Physik zu erstellen, für seine kritisch konstruktiven Anregungen und für seine motivierende Unterstützung, die zum Gelingen der Arbeit beigetragen haben.
- Frau Prof. Dr. Rita Wodzinski sowie Herrn Prof. Dr. Lutz-Helmut Schön für die Übernahme der weiteren Gutachten.
- den Mitgliedern der Arbeitsgruppe für ihre Unterstützung und für die freundliche Arbeitsatmosphäre. Mein besonderer Dank innerhalb der Arbeitsgruppe gilt Herrn Thomas Schmidt für seine stete Bereitschaft, mir bei technischen Herausforderungen hilfreich zur Seite zu stehen.
- den Schülerinnen und Schülern sowie den betreuenden Lehrerinnen und Lehrern für ihr Interesse und ihre Bereitschaft, an der Studie mitzuwirken.
- den studentischen Mitarbeiterinnen und Mitarbeitern für die gewissenhafte Aufnahme und Codierung der Beobachtungsdaten.
- der Wolfgang-Seel-Stiftung für die Verleihung des Preises zur Förderung der Promotion von Lehrerinnen und Lehrern.

Ganz besonders danke ich meiner Frau Barbara sowie unseren Kindern Jonathan und Konstantin, die mich während der Zeit an dieser Arbeit getragen haben.

Erklärung

Versicherung gemäß § 7 Abs. 3 der PromO Nr. 17/2012

Hiermit erkläre ich, die Dissertation selbständig und nur unter Verwendung der angegebenen Hilfen und Hilfsmittel angefertigt zu haben.

Ich habe mich anderwärts nicht um einen Doktorgrad beworben und besitze einen entsprechenden Doktorgrad nicht.

Zudem erkläre ich die Kenntnisnahme der dem Verfahren zugrunde liegenden Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät I der Humboldt-Universität zu Berlin.

Düsseldorf, den 16. Oktober 2012

Stefan Kirchner

